UNIVERSITY OF SOUTHAMPTON

Faculty of Engineering and Physical Sciences
School of Electronics and Computer Science
Agents, Interaction and Complexity Group

Influence Maximisation and Adaptive Control of Opinion Dynamics on Complex Networks

by

Guillermo Romero Moreno
ORCiD: 0000-0002-0316-8306

A thesis for the degree of
Doctor of Philosophy

December 2023

University of Southampton

Abstract

Faculty of Engineering and Physical Sciences
School of Electronics and Computer Science

Doctor of Philosophy

Influence Maximisation and Adaptive Control of Opinion Dynamics on Complex Networks

by Guillermo Romero Moreno

Opinion shaping, by which a strategic attempts to influence the opinion of a social group, is a pervasive phenomenon in human behaviour, with clear examples in current societies in information campaigns, political competition, or marketing. However, the effect that influence attempts have in societies is not easy to foresee, as they are complex systems where interactions can cascade and compound in unpredictable ways. The research subfield of opinion dynamics employs a mathematical modelling approach to study these phenomena using tools from the more general field of complex systems. In this approach, the opinions held by an individual are typically modelled as mathematical variables, with changes in opinion dictated by simple rules triggering upon interactions, which typically only happen interact according to a complex network that reflects their social structure. By drawing from techniques from statistical physics, agent–based simulations, and optimisation research, this thesis studies the effects that different external control strategies have on opinion formation processes.

We first focus on a case where the external controller is a 'perfect optimiser', i.e. they can split their influence targets among the individuals and have information to do so strategically as to achieve the best possible result — a problem commonly known as *Influence Maximisation*. We place this optimising controller against an opponent in a scenario where individuals have preferences over two possible choices (e.g. products or political parties), propose an optimisation algorithm to find optimal targetting strategies, and analyse the characteristics of these to understand why they are effective. We find that optimal strategies can be characterised by two heuristics, *shadowing* and *shielding*, while the structure of the network only plays a secondary role. *Shadowing* entails targeting the same individuals as the opponent to directly block her influence in the network, unless the opponent's influencing power is much higher, in which case these are avoided. *Shielding* entails ring–fencing the individuals targeted by the opponent to indirectly block the spreading of her influence. We then

modify the previous scenario to incorporate different levels of bias against either opinion in individuals, and analyse optimal targetting strategies when the population is structured in different network topologies. We find a general pattern in which individuals that are difficult to control (i.e. biased against the controller, highly connected, or targeted by the opponent) are avoided if the influencing power is small and sought if the influencing power is high.

Last, we shift the scenario to one in which the controllers are not 'perfect optimisers' any more but only have very limited information and perform local moves to improve their situation in the short term. Therefore, their control strategies are 'adaptive', reacting to the dynamics of the opinions as they unfold and creating a dynamic interaction between the two. We focus on the specific agenda–setting scenario where political parties seek to increase votes by affecting the importance that different political dimensions have and how their strategies interferes with the process of arriving at consensus or polarisation within the social group. We find in this scenario that party competition often fosters the arrival of a polarised state with most individuals gathering in two opposed camps, although if parties perform frequent shifts in the issues they give importance to, their behaviour inadvertently fosters the arrival at a consensus in opinion in social group.

Contents

List of Figures

List of Tables

Declaration of Authorship

I declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. 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;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. 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;
5. I have acknowledged all main sources of help;
6. 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;

7. Parts of this work have been published as: Romero Moreno, G., Tran-Thanh, L., and Brede, M. (2020b). Continuous influence maximisation for the voter dynamics: is targeting high-degree nodes a good strategy? In *Proc. of the 19th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2020), Auckland, New Zealand, May 9–13, 2020*. IFAAMAS, 3 pages
Romero Moreno, G., Tran-Thanh, L., and Brede, M. (2020c). Shielding and Shadowing: A Tale of Two Strategies for Opinion Control in the Voting Dynamics. In *Complex Networks and Their Applications VIII. COMPLEX NETWORKS 2019*, volume 881, pages 682–693. Springer, Cham
Romero Moreno, G., Chakraborty, S., and Brede, M. (2021a). Shadowing and shielding: Effective heuristics for continuous influence maximisation in the voting dynamics. *PLOS ONE*
Romero Moreno, G., Manino, E., Tran-Thanh, L., and Brede, M. (2020a). Zealotry and Influence Maximization in the Voter Model: When to Target Partial Zealots? In *Springer Proceedings in Complexity*, pages 107–118
Romero Moreno, G., Padilla, J., and Brede, M. (2021b). The effects of party competition on consensus formation. In *3rd International Workshop on Agent-Based Modelling of Human Behaviour (ABMHuB'2021)*

Signed:..... Date:.....

Acknowledgements

The main thanks goes to my supervisor, Markus Brede, who has been of immense help of going through this journey. I am aware that not everyone is lucky enough to have such an attentive and generous supervisor, and much of the credits of reaching to the completion of this thesis go to him. I must also mention all the PhD students, post-docs, and faculty members from the AIC (Agents, Interaction, and Complexity), VLC (Vision, Learning, and Control), and WAIS (Web and Internet Science) research groups at the University of Southampton, with whom I have shared these beautiful years and many table football games. A space here is also reserved for my family and all my other friends in Southampton (and everywhere else!) for their support.

Chapter 1

Introduction

It has been argued that the evolutionary origin of reasoning is rooted in trying to persuade others to align their intentions (Norman, 2016); hence attempting to influence others' opinions may be as old as reasoning itself. The evolutionary benefit of persuasion comes from achieving common outlooks that can either improve the effectiveness of mutually beneficial collaboration or by obtaining an argumentative advantage at the expense of the persuaded partner.

As societies progressed, advances in communication technologies brought more powerful channels through which to exert influence, from the invention of the press to the radio and television. Massive diffusion of propaganda and advertisements marked much of the social opinion sphere in the last century. The recent arrival of the Internet has further revolutionised communication, moving from a centralised paradigm —where the channels are owned by a few powerful institutions— to a distributed one, where any individual can be a source of information and the propagation of opinions depends on individuals replicating them (Bakshy et al., 2012; Kramer et al., 2014; Quattrociocchi et al., 2015). Furthermore, the abundance and immediacy of virtual communication have led to an almost instantaneous spread of information (Bakshy et al., 2011; Morales et al., 2014).

Unexpectedly, the elimination of entry barriers brought by the so–called 'social media' — online platforms where anyone can share and broadcast messages and content at different levels — has not clearly democratised the opinion market. The topics under discussion and the prevalent views are still greatly controlled by powerful agents that continue to exert their influence on communication channels (De Domenico and Altmann, 2020; Morales et al., 2014; Borondo et al., 2012). Topical examples include the support campaign for Donald Trump (Badawy et al., 2018; McFaul and Kass, 2019), Russia's attempts to shape international opinion around the incidents in Ukraine[1], or

[1]Source: https://www.buzzfeednews.com/article/maxseddon/documents-show-how-russias-troll-army-hit-amer retrieved on October 4, 2023.

the Brexit campaign in the UK[2]. Such cases raise concerns about the threats influential agents pose for the stability of democratic systems in the new hyper–connected reality, demanding a better understanding of the dynamics caused by external influence on a social network (Hoferer et al., 2020; Stewart et al., 2019; Albanese et al., 2020).

Furthermore, the recent decentralisation of social communication introduces further complexity to the mechanisms through which opinions are formed in social systems or how collective decisions are made. This is because the amplification of continuous micro–interactions between individuals may lead to highly unexpected system–level dynamics (De Domenico and Altmann, 2020). As a consequence, a plethora of new opinion phenomena has arisen in post-Internet societies, such as echo chambers (Del Vicario et al., 2016; Choi et al., 2020), viral posts (Morales et al., 2014; Garcia et al., 2012), or opinion polarisation (Ramos et al., 2015a; Axelrod et al., 2021), finding much attention from both researchers and the broader public. Understanding these phenomena requires a complex systems approach that incorporates in its explanation the complexity of networks of interaction and non–linear effects (Biswas et al., 2018; Noorazar, 2020).

Understanding how an influence attempt interacts with opinion phenomena is not a trivial task, as, from the point of view of control theory, it adds a feedback interaction loop to an already complex scenario. Therefore, there are many questions that arise regarding the effects that an influence attempt may have. First of all, one may wonder if the external control is capable of affecting the dynamics of opinion and, naturally, if it is able to benefit from it. There may be circumstances where the means or resources that the external influence has are not enough for causing significant effect — or even more, they may result in the opposite effect that the desired one, what is usually known as 'backfire effects' (Carletti et al., 2006; Vendeville et al., 2022). Second, there is a question of the strategies that the external influence can employ and how they differ. Related to this, there has been quite some efforts in past literature to understand the problem termed *influence maximisation* (IM) (Domingos and Richardson, 2001; Kempe et al., 2003; Li et al., 2018), which addresses how to optimally target a limited number of individuals for attaining the highest influence possible in the social group. Third, multiple external controllers may be present and interacting in a quest to bring their desired opinions in the population. This introduces the question of how their influence effects combine — e.g. they may cancel out — and the suitability of different strategies to combat the other controllers (Bimpikis et al., 2016)s. Last, if the influence control reactively adapts to the dynamics of the opinions, there may be unexpected dynamics emerging, and it is important to understand how this interaction affects the outcome of the opinion process, such as if they increase a state of polarisation or fragmentation of opinions (González-Avella et al., 2006).

[2]Source: https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election retrieved on Ocober 4, 2023.

These questions are relevant in a wide number of contexts. First, these are very natural questions in the political realm, in the context of the promotion of candidates or parties in elections (Javarone, 2014; Wilder and Vorobeychik, 2018; Sobkowicz, 2016), favouring stances on policy issues (Moya et al., 2017), or agenda–setting strategies (De Sio and Weber, 2014). In this context, many scenarios typically involve multiple external controllers with opposing needs, so a game–theoretical study of how combinations of strategies perform is of particular interest (Fowler and Laver, 2008), as well as how their behaviour affects voters as a side effect (Fowler and Smirnov, 2005). As parties typically have limited information about the positions of the electorate, it is common to model their behaviour as adaptive, reacting to polls or outcomes of elections with simple rules (Kollman et al., 1992). The influence maximisation problem is also of paramount interest in marketing, closely related to 'viral marketing' or the use of 'influencers' in social platforms (Domingos and Richardson, 2001; Kempe et al., 2003). In this context, firms seek to spread the popularity of their product or have it preferred over those of the competition. These scenarios may include a single external controller if the goal is to make a new product known (Domingos and Richardson, 2001) or multiple external controllers competing to promote alternative products, adding a game–theoretical component (İkizler, 2019). Likewise, influence maximisation can also be applied to policy making with various purposes, as to promote cooperative behaviour (Han et al., 2018), prevent substance abuse (Rahmattalabi et al., 2018), increase awareness of risks (Yadav et al., 2017), promote the adoption of norms or ecological behaviour (Zhang et al., 2016), vaccine immunisation campaigns (Velásquez-Rojas and Vazquez, 2017; Wang et al., 2016), or limiting the spread of misinformation (Budak et al., 2011; Tsai et al., 2012). These scenarios typically include a single external controller (the policy maker) who wants to spread a piece of information effectively or needs to bring the social system to a desired equilibrium state. Last, influence maximisation also relates to other problems with similar frameworks but not strictly within opinion dynamics, such as guiding technological innovation (Alshamsi et al., 2018a) or preventing epidemic contagion (Altarelli et al., 2014; Dezső and Barabási, 2002; Fekom et al., 2019; Wilder et al., 2018b).

Although the study of external control has many applications and uses in different contexts, it is not free of ethical concerns. Improving understanding of how external control influences opinions, also indirectly provides tools to these external controllers to better do so. And while these attempts can be done to improve the welfare of the social group — e.g. all the scenarios provided above in the context of policy making, with examples of actual implementations in Zhang et al. (2016) and Yadav et al. (2017) — they can also be employed by malicious agents, authoritarian regimes, or firms, at the expense of the society. However, we argue that investigating these topics from public research institutions entails that this knowledge can remain publicly available to the wider society. Understanding the effects of external control not only provides tools for external controllers — with moral or immoral goals — to increase their

influence, but at the same time provides tools to the society to prevent this control, as there may be modifications in their micro-behaviours or in the patterns of social connections that may make the social group more robust to these influence attempts. Meanwhile, it can be assumed that similar research is also being undertaken privately by intelligence services and firms for their benefit. So an important question to answer is: when providing resources to better understand external control, are these increasing the tools that malicious external controllers already have more than the tools that benevolent external controllers or citizens have, or are these strengthening the latter group due to a current imbalance in resources? Either way, these key ethical questions have not found much discussion in the research around external control barring some exceptions (Hegselmann et al., 2015), and it is important that these concerns are voiced more often.

1.1 Research challenges

Understanding the effects of external influence on a social system is a task that faces important challenges, as opinion dynamics are highly complex processes that are hard to conceptualise, measure, and model. The typical conceptualisation of opinion formation processes involves a system of many agents that interact dynamically and who are typically embedded in a network of interactions with non-trivial topologies. Predicting system behaviour or analysing its characteristics may be challenging — particularly for non-linear interaction rules — while the parameter space can easily become high–dimensional, hindering its exploration or the optimisation of some outcome.

This high complexity is treated in the literature via two paradigmatic approaches (Jędrzejewski and Sznajd-Weron, 2019): *statistical physics* and *agent–based modelling*, both of which aim to extract simple, general patterns at the system level from the low–level behaviour of individual agents (Castellano et al., 2009a). In statistical physics, this is typically done by making certain assumptions of the low–level distributions and working with average behaviours and simplifications that may allow for analytical solutions. This approach is highly related to *sociodynamics* (Helbing and Weidlich, 1995), where social phenomena are studied from the perspective of dynamical systems and non-linear dynamics, typically modelling the rates of change in the system. This approach strongly relies on the tractability of the proposed models, so its success depends on selecting simple models and adequate approximations that preserve the main mechanisms of the modelled phenomena. In contrast, agent–based models rely on numerical simulations to explore the characteristics of the system behaviour, so they are not as strongly constrained by the model complexity and accept more details in the low–level interaction rules. This bottom–up approach has special relevance in the study of conditions that trigger the

emergence of system–level behaviours. While simpler models can afford to exhaustively explore large areas of the parameter space, models with wealth in details can rely on calibrating their parameters with empirical data. Due to the potentially complex models, care needs to be put in their design to cater for reasonable time and memory constraints in computing. Although both approaches are able to provide some insights into the studied phenomena, major challenges still remain in the analysis and understanding of the models, as they include a large number of variables that vary in time and that are embedded in complex networks structures.

Furthermore, performing influence maximisation or any other optimisation goal on opinion dynamics processes takes tractability challenges one step further. As evidence, the problem of influence maximisation has been proven to be NP-hard in its most common formulations in the literature (Kempe et al., 2003), resulting in many researchers focusing on devising heuristics or approximation algorithms to solve this problem while scaling well with system size (Chen et al., 2013; Li et al., 2018). Any influence maximisation attempt should devise an optimisation algorithm that provides an exact solution or bounds to an approximate solution (Liu et al., 2010; Mai and Abed, 2019). Although a significant body of literature has been devoted to devising algorithms that provide solutions, a further challenge remains in the examinations of the structure of these solutions, uncovering the ingredients that are responsible for them to be successful. This is an open question that, barren some exceptions (Yildiz et al., 2013; Lynn and Lee, 2016) has been largely unexplored.

Additionally, contexts of opinion dynamics often include more than one external controller willing to introduce their influence in a population —e.g. competing political parties aiming to collect partisans or companies that try to have their products preferred over the competition. Such scenarios add a game–theoretical component (Hegselmann and Krause, 2015), with research works typically focusing on finding Nash equilibria (Masucci and Silva, 2014) — i.e. combined strategies from where no player has an incentive to divert — or best responses to certain strategies held by the opponent (Javarone, 2014; Bharathi et al., 2007; Lin et al., 2015). The interaction of opposing controllers has been significantly less explored than when a single external controller is present and more research is needed to understand these scenarios.

Last, the presence of adaptive control that reacts to the opinion dynamics adds a challenge to its modelling an analysis, as the time scales in which the dynamical processes evolve and react to each other can play an important role in the outcomes. These two components are related by a feedback loop with high non-linearities and disentangling the mechanisms that cause specific outcomes can be challenging. Therefore, there is a need to better understand how adaptive controllers interfere with the opinion dynamics and lead to new emerging phenomena (Fowler and Smirnov, 2005; González-Avella et al., 2006).

1.2 Research contribution

This work studies different scenarios of external influence on opinion dynamic processes on graphs, emulating social systems. The approach is chiefly theoretical: we study system behaviours and optimal solutions, as well as their dependence on various aspects of the models, such as network topology, model parameters, or influence power. By doing this, we aim to better understand the mathematical properties of dynamical social systems and their interplay with external sources of influence.

An important part of this thesis focuses on the case of external controllers as 'perfect optimisers' who have complete information about the social system and where opposing controllers compete. In this context, we focus on a scenario where external controllers exert influence by targetting agents in a population that is embedded into a social network, and allow them to divide their influence power continuously among all agents in the population. For this scenario, we first the infuence maximisation problem by developing an gradient–ascent algorithm that is able to arrive at optimal strategies of targetting with high guarantees and in polynomial time. This is novel in that most previous work aims to find an optimal subset of nodes to influence with a fixed targeting intensity, while our modelling adds flexibility in targeting, allowing for arbitrary intensity (under a certain budget).

Additonally, and unlike most previous research in the field, we perform a thorugh analysis of the structure of optimal targetting strategies to characterise the factors that shape their structures, to provide insights of what aspects of the strategies make them effective and how these change under different scenarios of power imbalance or react to the strategy of the opponent. To complement this analysis, we also derive mean–field approximations that provide further insights between the mathematical relations of the strategies to other system variables under specific limiting conditions.

As an important contribution, we distil two joint heuristics that can explain a large part of the variance in optimal influence allocations, which we term *shadowing* and *shielding*, respectively. With shadowing, an external controller focuses resources on nodes targeted by the opponent if in resource superiority, while avoiding them when in resource disadvantage. With shielding, the controller focuses additional resources on nodes surrounding those targeted by the opponent, thus preventing the spread of influence, although an inverse effect may also appear if the controller's budget is much lower than the opponent's. Interestingly, these two heuristics strongly depend on the behaviour of the opponent, while the effect of the topology of the social network —e.g. node degree or centrality, which are common heuristics in the literature— is only marginal and purely a side effect resulting from the application of shadowing and shielding.

While we initially explore the relevance of these heuristics in the regime of continuous influence maximisation (where nodes can be targeted with arbitrary intensity), we also test these heuristics for the discrete influence framework of homogenous targeting intensity, which is more common in the influence maximisation literature (Li et al., 2018; Peng et al., 2018). We observe that shadowing and shielding are also present in solutions for the discrete framework, although their presence is also shared with an intrinsic dependence on node degree, revealing an important difference between the continuous and discrete frameworks.

Additionally, we also explore influence maximisation on heterogeneous populations where some individuals are biased towards one of the opinions. This additional aspect accounts for the natural tendencies of individuals towards certain preferences or, alternatively, the presence of external factors not captured in the model. We address whether these biased individuals are worth being targeted, under which circumstances, and how much effort should be spent on them. We find that targeting strongly biased agents is only optimal when the controller has an abundance of influence resources, while they should be avoided otherwise. Interestingly, the topological position of biased (and unbiased) agents is highly important in these scenarios, with hubs also being preferred when influencing budgets are high.

The last section of this thesis moves away from the assumption that external controllers have full knowledge of the opinion dynamics and that they can predict its evolution in the long term. Instead, we assume that they adaptively react to the current state the opinion dynamics and their opponents at similar timescales, and our focus is on how the presence of this adaptive control interferes with the trajectory of the social dynamics. In particular, we consider the scenario of political parties that affect a population's consensus formation process by modifying the importance in which political dimensions are perceived by the population. Unlike the above, we consider here the case of opinions that have real values from within a continuous interval, representing political dimensions of opinions. Although there are numerous works studying the adaptive control of parties in the political sciences, its interference with the opinion dynamics and their feedback effect have not been previously studied. By employing agent–based simulations, we show that the adaptive control by the parties tends to bring the citizen population to a state of polarisation that would have not happened without their interference, although under some circumstances their adaptive control can also foster the arrival at a population–level consensus. The produced effect depends on the constellation of party positions and their number, although a polarising effect is much more prevalent than a promotion of consensus. If the population is embedded into a sparse, random social network, it is more prone to reach a consensus state, but the effects that party competition has on this process remain qualitatively similar.

1.3 Thesis structure

The structure of the thesis is as follows. First, Chapter 2 provides background theory of the different areas that overlap within the topic of this thesis. This chapter includes an introduction to how social networks are commonly model and some tools in network science for their analysis, presents the most important modelling approaches to opinion dynamics, covers relevant literature on external influence in opinion dynamics, introduces famous agent–based modelling for social simulations, and covers the most important techniques to perform optimisation. Then, Chapter 3 focuses on the first scenario where the external controller optimises its influence strategy against an opposing external controller. There, we formalise the continuous influence maximisation problem with arbitrary targeting intensities, provides a gradient–ascent algorithm for obtaining optimal solutions and analyses the structure of these solutions in relation to the strategy of the opponent and the role of individuals within a complex networks. Chapter 4 extends the scenario and analysis to the case of heterogeneous populations with biased agents, providing understanding of optimal influence in this richer scenario. Chapter 5 moves away from external controllers as 'perfect optimisers' and investigates the case of adaptive external control in the context of party competition affecting the agenda–setting and their effect in consensus–formation processes within a society. Last, Chapter 6 summarises the contributions of the thesis, performs a critical review of its limitations, and sketches possible lines of investigation that can extend the reach of the current work.

Chapter 2

Background theory and literature review

Perhaps motivated by the increasing prevalence of social media and their influence on public opinion, processes of opinion formation on social networks have found much attention in the recent literature (Biswas et al., 2018; Condie and Condie, 2021; Lim and Bentley, 2022). Models in this domain have addressed general properties of opinion spread on static and co-evolving networks, but also basic mechanisms underlying phenomena such as radicalisation (Ramos et al., 2015b; Lim and Bentley, 2022) and the role of external influence (De et al., 2018; Palombi et al., 2017; Macy et al., 2021).

Opinion dynamics have been commonly studied through a statistical physics approach. This is not surprising, as individuals in a social system can be paralleled to particles with local rules that lead to various system–level behaviours (Castellano et al., 2009a). The study of opinion formation from this perspective can help to determine which conditions may drive a society to a desirable state, which can be derived either analytically or via simulations.

When modelling opinion dynamics, there are three main aspects upon which researchers must decide: how to define influence ties between individuals, mathematically model opinions, and how their opinions change based on those influencing ties. The first aspect is addressed in Section 2.1 of this chapter, which introduces network formalisations of social systems and their typical characteristics and structures. The second and third aspects are jointly covered in Section 2.2, which reviews the basic ideas behind the main opinion formation models. Section 2.3 covers an additional ingredient to opinion dynamics that is the focus of this thesis: external influence on opinion formation, covering the recent literature on both untargeted and targeted external influence. Section 2.4 reviews Agent–Based Modelling as an important tool to study social systems, covering its history and relevant applications to social and political phenomena. Section 2.5 introduces the topic of optimisation and

the basic methods for performing it. Finally, Section 2.6 summarises the chapter and points to existing gaps in the literature and the fit of our work.

2.1 Network models of social influence

A society is a complex system formed by individuals that have frequent interactions. However, individuals typically only interact with a small subset of others due to spatial proximity or social ties, creating a complex landscape of interconnections (Guimerà et al., 2003). A common and helpful approach to understand the structure of interactions is to use techniques from graph theory and model them as a network, where individuals are nodes and interactions are represented by edges (Newman, 2003).

Formally, a graph (or network) $G = \{V, E\}$ represents a social group, where nodes $v_i \in V, i = 1, \ldots, N$ correspond to the individuals and edges $e_{ij} \in E$ that interactions can occur between the pair of nodes v_i and v_j (Majeed and Rauf, 2020). In many cases, it is useful to assign a weight $w_{ij} \in \mathbb{R}^{0+}$ to each edge e_{ij} to reflect the strength of the interaction (e.g. frequency, duration, effect) which makes the graph *weighted*. Under this formulation, an absence of an edge can be modelled as weight with $w_{ij} = 0$, and *unweighted* graphs would correspond to weights with binary values, $w_{ij} \in \{0, 1\}$. All pairwise weights can then be gathered into a *weighted adjacency matrix* $W = (w_{ij})$, which fully captures the network (Newman, 2003). Another particularly useful quantity is the degree $d_i = \sum_j w_{ij}$ of a node, as it provides insight of how 'socially active' an individual is[1]. Further, inspecting the distribution of degrees of a network provides a very valuable insight on its structure and the level of heterogeneity in nodes' connection patterns. The weighted adjacency matrix and node degrees can be merged into the *Laplacian matrix* of the graph, $L = D - W$ — where D is a diagonal matrix whose elements in the diagonal correspond to node degrees, i.e. $D_{ii} = d_i$ (Newman, 2003). Thanks to the mathematical properties of this matrix, many powerful tools from linear algebra can be employed to unveil structural properties of the network, such as the presence of clusters (Newman, 2003).

There are a few useful measures that characterise the role a node has in the network topology. The *geodesic distance* between two nodes reflects the minimum number of links required to travel between them — i.e. the length of the shortest path that joins them. The *centrality* of a node relates to its global position within the graph and it comes in different flavours. The *betweenness centrality* is the fraction of shortest paths in the network that pass through a node and reflects the ability of a node to act as an intermediary (Newman, 2003). The *closeness* centrality is the inverse of the sum of

[1]Some works use the term '*degree*' only with unweighted networks, while they term the sum of weights as '*strength*' in weighted networks. We will use 'degree' in both cases for simplicity of language, as the distinction is irrelevant in our work.

shortest distances to all other nodes in the network and reflects how quickly information can spread from a node to the rest of the network. Another important centrality measure is the *eigenvector centrality*, which is given by the eigenvector of the largest eigenvalue of the Laplacian matrix, and reflects the idea that nodes connected to 'important' nodes should also be important. Variations of the eigenvector centrality are the *Katz centrality* or Google's *PageRank*, which was designed to rate the relevance of websites for ranking them in its search engine (Majeed and Rauf, 2020). Note that many of these centrality measures are only meaningful if the network is *connected*, i.e. at least one path connecting each node to every other node exists.

There are many extensions of the basic graph formulation above adding ingredients that capture further details of the system that is modelled. An important distinction is in whether edges have a *direction*, implying that the interaction has an actor and a receiver. This modification results in an adjacency matrix that is not necessarily symmetric any more, so some of the metrics introduced above need to be redefined. For instance, the degree of a node needs to be divided into two different types of degree, the *out-degree* $d_i^{in} = \sum_j w_{ij}$ and the *in-degree* $d_i^{out} = \sum_j w_{ji}$, each of them defining different types of roles of nodes in the network (Newman, 2003). Other relevant extensions include *multi-layer networks* (Kivela et al., 2014) — where each layer represents different types of interactions, with some form of coupling between the layers — networks with higher–order interactions (Battiston et al., 2020) — where each edge links three or more nodes simultaneously — temporal networks — with time–stamped interactions — or dynamic network — whose structure changes with time, maybe coupled to the social interactions.

Since this work focuses on opinion dynamics, we will use networks to represent channels of influence among individuals in their opinion formation. Ideally, we would employ empirical networks obtained from data regarding influence channels, as to represent a credible social influence structure. However, the task of registering influence among individuals can be problematic, since it cannot be directly measured. Instead, it needs to be inferred from recurrent data over long periods of time and at society–wide levels, which is often very difficult to access. Furthermore, influence may vary dynamically, so the confidence over its inference cannot be reliably increased by extending the time scope of measurements. Due to all these difficulties, measuring and constructing influence networks requires a whole branch of research and it is still in its infancy (Peng et al., 2018; Panzarasa et al., 2009).

As simpler alternatives, influence networks can be proxied by registered exchanges of information, such as in-person conversations (Mastrandrea et al., 2015), 'follow' patterns on social media (Atienza-Barthelemy et al., 2019), or email exchanges (Guimerà et al., 2003). Although these proxy networks are not identical to the influence network we are after, their structure and characteristics may resemble the target ones well enough to produce similar outcomes in the opinion dynamics.

Depending on the context that is studied, these networks could be undirected (if they represent friendship, family relations, or bilateral conversation) or directed (if they capture the emission of messages with a sender and a receiver, such as in some social platforms or in emails). Additionally, it is also common to build synthetic network topologies that capture basic network features of social networks. Since they are synthetically created, specific network properties can be controlled to better understand their effect on the social dynamics. Below we review the most prominent synthetic network architectures in the social network literature.

2.1.1 Complete graph

The complete graph is a graph where every node is connected to any other node in the network, which would for instance correspond to a closed (and typically small) group of people where everyone interacts with each other. If the network is unweighted, this implies that the there is a lack of social structure, serving as a baseline to decouple its effect to that of the opinion formation process.

2.1.2 Bipartite graph

A bipartite graph is a network with two classes of nodes, A and B, and where connections between nodes of the same class are forbidden (Newman, 2003). These can be used to represent e.g. interactions between buyers and sellers or ties between workers and organisations. They can also be used to connect individuals as one type of entity to interactions, representing the other type of entities in the network, and thus capture higher–order relationships using pairwise edge connections (Battiston et al., 2020).

2.1.3 Spatial networks

Spatial networks refer to networks where nodes are embedded into a geographical (typically Euclidean) space and are linked by edges depending on proximity. These networks capture interactions that are strongly embedded geographically (such as in supply networks) or in an abstract 'social space' (Hamill and Gilbert, 2009). The presence of edges can be deterministic (if the distance between two nodes is smaller than a threshold) or probabilistic (with a function depending on the distance). A subclass of spatial networks that are particularly used in simulations of social systems are *regular networks*, which have the benefits of high tractability and easy visualisation (Amblard et al., 2015). These are typically formed by square lattices, influenced by the early research in cellular automata (Amblard et al., 2015), although can contain other regular patterns, and span different numbers of dimensions. They may be infinite or

may have aperiodic or periodic boundaries, with the latter implying that nodes at one boundary are connected to the nodes at diametrically opposed boundary.

2.1.4 Erdős–Rényi random graphs

Erdős–Rényi random graphs (Erdös and Rényi, 1959) are among the most common synthetic network structure used in research involving social simulation (Amblard et al., 2015). Edges in Erdős–Rényi random graphs are sampled from a Bernoulli distribution with equal probability η, resulting in a node degree distribution following a binomial distribution. This results in an homogeneity that them share many of the properties held by complete networks while having a sparsity that can be easily controlled with the parameter η. However, values of η below $ln\, N/N$ (with N being the size of the network) can arrive at fragmented networks with many disconnected groups of nodes, so it is usually undesirable to use value below that limit, known as the *percolation threshold* (Erdös and Rényi, 1959).

2.1.5 Small–world networks

Small–world networks (Watts and Strogatz, 1998) are a class of networks that showcases some of the properties found in human social networks. They were inspired by an experiment suggesting that the average number of connections needed to link any two individuals in a society is around six (Milgram, 1967; Guare, 1990). Small–world networks can be seen as an interpolation (controlled by the parameter $\gamma \in (0,1)$) between regular lattices and *Erdős–Rényi* random graphs, as they maintain the small mean geodesic distance of random graphs [2] —i.e. any node is easily reached by any other node in the network — and the high clustering coefficient of lattice graphs —i.e. neighbours of a node also tend to be neighbours themselves. The standard algorithm for building small–world networks departs from a ring of nodes where each node is connected to its k closest neighbours and then each edge has one end rewired to a new random node with probability γ (Watts and Strogatz, 1998) [3], therefore mixing the mechanisms to generate regular lattices and Erdős–Rényi random graphs. Small–world networks fit well structures of societies under loose spatial restrictions (Newman, 2003), as happens with in-person communications, although some of their characteristics may also be found in online communities (Panzarasa et al., 2009).

[2] More precisely, a mean geodesic distance that scales logarithmically with the network size (Newman, 2003).

[3] Duplicated edges or disconnected graphs are typically not allowed.

2.1.6 Scale–free networks

Scale–free networks are a class of networks that have received much attention, not only for their pervasiveness in social systems (Broido and Clauset, 2019) but also in many other domains (Barabási and Albert, 1999). They are characterised by a scale–free (power law) distribution in the node degree, i.e.

$$P(d_i) \propto d_i^{-\alpha}, \tag{2.1}$$

where $P(d_i)$ is the probability of finding a node of degree d_i and α is a constant value that can vary between networks. Consequently, a few nodes are extremely well–connected, while the vast majority have very few connections, showcasing a heavy–tailed distributions of node degrees.

It is hypothesised that these network structures commonly originate from rich–get–richer network formation dynamics Barabási and Albert (1999), where new nodes attach to existing nodes with a probability proportional to their degree. Networks generated by this process are commonly referred to as *Barabasi–Albert networks* (or BA–networks, for short), named after the authors of the seminal paper Barabási and Albert (1999). BA–networks are equivalent to scale–free networks with $\alpha \in (2, 3]$ [4]. Another standard algorithm for generating general scale–free networks is the *configuration mode*l (Bayati et al., 2010), which generates random networks from a pre-defined degree distribution sampled from a power–law distribution.

2.2 Modelling opinion formation processes within a society

The psychological foundations of why opinions spread through a society can be found in two basic mechanisms: the *social impact theory* (Latané, 1981) and the *persuasive argument theory* (Vinokur and Burstein, 1974; Myers, 1978). Although stemming from different mechanisms —the need to fit in a group or an active exchange of arguments, respectively— both theories account for the same behaviour: individuals tend to change their opinion and align towards their peers'. This is the basic ingredient in all opinion dynamics models; implementation details, however, vary widely in their specification of opinion change upon interaction. This is natural since the phenomenon that is to be modelled (the changing of individuals' opinions based on their peers' influence) is highly complex. Furthermore, it is not easy to harvest empirical data that can be used for model validation, and models commonly lack links to sociological studies (Flache et al., 2017). This results in a plethora of different

[4] This range of α values give scale–free networks interesting properties, such as having a first moment (which would not happen with $\alpha < 2$) but not a second moment (which emerges with $\alpha > 3$). However, note that social networks do not always fulfil these characteristics (Broido and Clauset, 2019).

models that are considered in the field, often varying significantly in their dynamical properties. The decision on which one to use depends on the social process to be modelled or, sometimes more pragmatically, on the need for mathematical tractability.

Opinions are typically transferred to mathematical quantities as a variable or set of variables. The simplest models include a single binary variable — e.g. being in favour or against an idea (Granovetter, 1978) — which can be naturally extended to a discrete variable — e.g. choosing among competing products or political parties (Marvel et al., 2012). Less coarse–grained representations are adopted with continuous variables, accounting for degrees of intensity in which a particular position is embraced (Hegselmann and Krause, 2015). More ambitious models are multi-dimensional, with each individual holding several independent opinion variables associated with each of the dimensions, which can at times also be binary (Mueller and Tan, 2018), continuous (Fortunato et al., 2005; Schweighofer et al., 2020a), or a combination of both (Martins, 2008).

A critical reader can note that there is a considerable simplification when reducing sophisticated mental phenomena to one or few variables. While this is true, the aforementioned opinion models attempt to capture the most relevant opinion states for answering specific research questions, while other more superfluous aspects can be included in the form of noise (Brede et al., 2019b; Castellano et al., 2009a). Furthermore, typical research questions answered by these models do not focus on understanding specific events or predicting the behaviour of particular societies but attempt to understand general behaviours —or *stylised facts* (Schweighofer et al., 2020a)— that emerge from different interaction rules between individuals. Hence, most research in the field does not employ real opinion data (whose mining and distillation into abstract opinion variables is challenging) but makes assumptions on general opinion distributions (Flache et al., 2017).

Most relevant models in the literature are introduced next. We have categorised into two main groups depending on whether opinions are modelled as discrete or continuous variables. For more details of the models and most extensive reviews of opinion dynamics, refer to Castellano et al. (2009a); Sîrbu et al. (2017); Proskurnikov and Tempo (2017, 2018) and Grabisch and Rusinowska (2020).

2.2.1 Models considering opinions as discrete variables

In discrete–variable models, there are only a few possible opinion states, and changes in opinion are consequently abrupt. Although this may appear as a limitation to the model richness, it is argued that humans are not psychologically comfortable at perceiving degrees between choices and rather think in symbolic streams (Mueller and Tan, 2018; Rabinowitz and Macdonald, 1989). In these models, influence is

implemented as *copying* opinions from peers, and models mainly differ on which opinions are copied, in which order, or with which probability.

For instance, in the paradigmatic *voting dynamics* or *Voter Model* (Clifford and Sudbury, 1973) individuals are randomly selected to then copy the opinion of a random neighbour, entailing high levels of stochasticity. The probability of copying either opinion depends on how many neighbours are holding that opinion, which is a linear relation that makes the model highly tractable. The *majority rule* reduces the stochastic element by having individuals adopting the opinion of the majority of their neighbours, although the selection of which individuals performs the copying each time is also asynchronous and random (Galam, 2002). This can be considered of model incorporating *complex contagion* in that the aggregation of the neighbouring opinions is highly non-linear (a step function). Majority rules drive the social group to a consensus much faster than in the voter model (Valentini et al., 2016). The Glauber dynamics of the *Ising model* (Ising, 1925) likens opinion diffusion to the coupling of spins in ferromagnetic materials and can be thought of as an in-between the Voter Model and the majority rule, as the probability of copying an opinion increases super-linearly with its prevalence on an individual's neighbourhood (Galam et al., 1982). An additional and important ingredient is the inclusion of random changes in spin that are controlled by a 'temperature' parameter T.

With a similar behaviour to the Glauber dynamics, *social impact* models (Nowak et al., 1990) draw inspiration in the social impact theory (Latané, 1981) to include heterogeneity in the ability to convince individuals with the same opinion ('supportiveness') and or the opposite opinion ('persuasiveness'), plus a 'social distance' structure. When simulated in a square lattice, the model shows the appearance of clusters of opinions (Nowak et al., 1990), although when noise is present the opinion show a 'staircase' dynamics by which clusters shrink between periods of stability until the population eventually arrives at a consensus (HOŁYST et al., 2001). Variations of the model include reinforcement dynamics in the supportiveness and persuasiveness (Kohring, 1996), the presence of a strong leader (Kacperski and Holyst, 2000), or a social temperature (the general rate of random behaviours) that is coupled with the dynamics (HOŁYST et al., 2001).

Other models are inspired by applications in other fields. For instance, the epidemic–related *SIR* (Susceptible–Infected–Recovered) (Kermack and McKendrick, 1927), which models the spread of a contagious disease, can be naturally mapped to attention mechanisms in phenomena like the spread of rumours, where 'infected' would imply an awareness of the rumour and the capability of spread it further and 'recovered' that the attention towards the rumour is lost. The variant *SIS* (Susceptible–Infected–Susceptible) would model a possible re-ignition of the contagious state. The *Naming Game* (Steels, 1995) was developed for the study of language dynamics and fixation of new works, where agents hold a pool of terms to

name a concept and perform one of two actions upon interaction: they copy one term from the other agent's pool if there is no common term between the two or if there is a common term in both pools they discard all others. As each term can be presence or absence, this opinion model can be conceived as a multidimensional binary model with particular rules.

Although all these models usually include binary variables, they can often be easily extended to categorical variables with more than two choices (Axelrod, 1997b) or with an in-between neutral state (Arendt and Blaha, 2015). In this class, the *Axelrod model* (Axelrod, 1997b), which includes multiple categorical variables, stands out in the study of cultural dissemination, drawing attention from both social scientists and physicists. In this model, when two agents encounter their probability of interaction is proportional to the overlap between their variables, hence encoding the notion of *homophily*, i.e. interactions occur more frequently between the alike. Unlike many of the models above, this model does not necessarily lead to a single cluster or consensus, but is capable to create under some parameter regimes an entrenched state of diversity into different cultural groups that do not interact with each other, reproducing behaviours commonly found in real societies.

Progressive (or *static*) opinion dynamics models form a sub-class of binary models that is particularly relevant in the influence maximisation literature from the computer science community. These models only allow changes of opinion in one direction, from *inactive* to *active* (Kempe et al., 2003). The system is typically initialised with a few active agents and many inactive ones, and the propagation occurs as a one–off process, with the new opinion percolating through the society. This modelling feature is particularly useful for situations when a new opinion is introduced into a social group (e.g. spread of rumours) and for opinions that remain unchanged for a long time once individuals commit to them (e.g. adhering to a telecommunications supplier, buying a product, (Leskovec et al., 2007)), but are inappropriate when addressing fast–changing opinions, such as stances toward political issues, or when longer time frames are studied. In the paradigmatic independent cascade model (IC) (Goldenberg et al., 2001), active agents can activate their inactive neighbours in a single attempt with a probability encoded by the weight of their shared edge. Once all active nodes have used all their edge attempts, the spreading process ends. In the related *linear threshold model* (LT) (Granovetter, 1978), inactive agents only change their state when the combined weight of incoming edges from active neighbours surpasses a specific threshold, similar to perceptrons in neural networks. Extensions to these models abound, including the addition of non-linear diffusion functions, time–varying parameters, or considering community structure, to name a few (Xie et al., 2021).

Next, the details and variations of the Voter Model will be presented, as it will be extensively used in this work.

The voter model

Most of this work employs the Voter Model (VM) (Clifford and Sudbury, 1973; Holley and Liggett, 1975) as the diffusion rule of opinions, since it is a well–studied model that has been prominent in the opinion dynamics literature for many years. In the Voter Model, opinions are typically binary —e.g. A or B, Conservative or Labour, etc.— although it can easily be extended to more than two states (Marvel et al., 2012). The diffusion rules are simple: at each time step, one agent is randomly selected and copies the opinion of a random neighbour. For the study of its dynamics, it is useful to consider the evolution of the probability of a node $x_i(t) \in [0,1]$ to be in state A at time t (Mobilia et al., 2007; Even-Dar and Shapira, 2011). From this perspective, the flow of probabilities in any node is described by the rate equation

$$\frac{dx_i}{dt} = (1 - x_i)\frac{\sum_j w_{ij} x_j}{d_i} - x_i \frac{\sum_j w_{ij}(1 - x_j)}{d_i} \,. \tag{2.2}$$

Note that the first term on the right–hand side of (2.2) represents the probability of holding opinion B and choosing a neighbour with opinion A (inflow of probabilities), and the second term corresponds to holding opinion A and choosing a neighbour with opinion B (outflow of probabilities). Importantly, the number of agreeing neighbours linearly affects the rate of change, which facilitates the tractability of the problem. This linearity in the influence of neighbours is a key feature of the VM, as it facilitates establishing many of its properties via mathematical derivations (Redner, 2019). Additionally, note that this probability flow can be equalled to continuous opinions and (2.2) rewritten as

$$\frac{dx_i}{dt} = \sum_j w_{ij}/d_i (x_j - x_i) \,, \tag{2.3}$$

which is analogous to the *DeGroot model*, a paradigmatic model of opinion dynamics with continuous opinions.

Although (2.2) implies that the expected probability of any node to be in state A converges to a single value $x_i = x$, the full probability distribution is doubled–peaked, with two absorbing states placed at the extremes of all agents holding the same opinion (A or B). Since opinions continuously flicker back and forth, homogeneous populations in finite, connected networks asymptotically reach consensus at one of these two absorbing states; which of them depends on the initial distribution of opinions, the network topology, and the stochasticity of the dynamics (Even-Dar and Shapira, 2011). Times to reach consensus vary significantly depending on the network topology, with sparseness in the number of links considerably delaying the process (Redner, 2019).

Nevertheless, the behaviour of the system changes significantly in the presence of agents with fixed opinions, commonly known as *zealots* (Mobilia, 2003), *stubborn agents* (Kuhlman et al., 2013), or *frozen nodes* (Chinellato et al., 2015). If zealots support only one of the opinions, the system always asymptotically converges to their opinion regardless of initial conditions. When there are zealots from both sides, the system converges to a dynamical, fragmented steady state with both opinions present (Mobilia et al., 2007) in proportions that depend on the topology of the network and the position of the zealots in it (Masuda, 2015). Importantly, no absorbing states exist for this case and the asymptotic probability distribution is singled–peaked and with non-negligible variance (Yildiz et al., 2013). Models of voting dynamics including zealots have been employed to model election outcomes. For example, Braha and de Aguiar (2017) have found voting dynamics with zealots to be suitable for modelling outcomes of USA elections in the last century, Vendeville et al. (2021) have used a similar approach for election outcomes in the UK, and Fernández-Gracia et al. (2014) have also applied noisy voter dynamics to model election outcomes in the USA. Additionally, note that the probability flow formulation of the Voter Model with zealots can be mapped to the Friedkin–Johnsen model of continuous opinions and private convictions under some formulations, or to a DeGroot model with zealots.

2.2.2 Models considering opinions as continuous variables

From the side of opinion models with continuous opinion variables, the *DeGroot model* (DeGroot, 1974) stands out for its simplicity, as agents just adopt a weighted averaged opinion of their neighbourhood (including themselves). However, such a simple model is considered unrealistic, as it always leads to a consensus of the population as long as the graph is connected (Mueller and Tan, 2018), which is at odds with the fragmented or polarised opinion landscapes that are the norm in many societies (Iyengar et al., 2019).

After the DeGroot model, a new generation of opinion models added mechanisms to allow for the emergence of polarised or fragmented opinion groups in the steady state. In the paradigmatic *bounded–confidence model* (BC) (Deffuant et al., 2000; Hegselmann and Krause, 2002)), interactions have no effect if agents' opinions are too different, which effectively breaks the social network into disconnected components (Kurahashi-Nakamura et al., 2016). The Friedkin–Johnsen model (Friedkin and Johnsen, 1990) adds fixed, private opinions to each agent, which affect their opinion update and may prevent the arrival to consensus in some regimes. These private opinions can also be mathematically equivalent to having zealots with fixed opinions in the society (Taylor, 1968). Other modifications of the DeGroot model that break the consensus outcome include repulsive forces (Jager and Amblard, 2005), or contrarian agents that seek to differentiate themselves from others (Mäs et al., 2010).

Inspired by the dynamics of rarefied gases, *kinetic models* of opinion formation (Toscani, 2006) build on the DeGroot model by introducing a noisy diffusion of the opinions and non-linearities, as both the rate of concession and the diffusion can depend on the value of the opinion. Drawing techniques from physics, solutions were found by operating on differential equations instead of performing simulation (Toscani, 2006). These modifications can still lead the population to a consensus around a single opinion, but some regimes also retrieve bimodal distributions in the steady state (Toscani, 2006; Lallouache et al., 2010; Sen, 2011; Oestereich et al., 2020). Next, the details and variations of the bounded–confidence model are presented.

Bounded confidence models

Chapter 5 employs the bounded–confidence (BC) model, as it is common in the study of consensus–forming and polarisation processes in societies (Ramos et al., 2015a; Deffuant et al., 2002). This model departs from the DeGroot model of opinion dynamics and introduces a limitation to interactions between individuals, which only have an effect if their opinions are close enough. This interaction rule requires the definition of a distance metric between opinions, and a *tolerance threshold* δ that determines the maximum distance for which interactions are still effective. The bounded–confidence mechanism is inspired by the idea that individuals whose opinions are *too different* may never listen to —or, even more, convince— each other. This might be related to the rise of polarisation in several Western countries, in which the attitudes and opinions towards those that are ideologically different have worsened over the last years (Iyengar et al., 2019). From the seminal studies of the BC model, it was found that the value of the tolerance threshold δ crucially influences the state to which the system will converge (Weisbuch et al., 2003), with outcomes ranging from consensus (all agents gathering into a single cluster with very similar opinions) to polarisation (two clusters of opinions) or fragmentation (many scattered clusters) (Sikder et al., 2020).

The BC model was developed in parallel by Deffuant et al. (2000) and Hegselmann and Krause (2002), leading to two fundamental variants, which will be labelled as DW and HK, respectively. In the DW model, two randomly chosen agents i and j encounter at each time step t and they only interact provided that the distance between their opinions, $o_i^{(t)}, o_j^{(t)}$, is less than the specified tolerance threshold δ. For a 1D opinion model, updates at each time step occur as

$$o_i^{(t+1)} = \begin{cases} (1-\mu)\, o_i^{(t)} + \mu\, o_j^{(t)} & \text{if } |o_i^{(t)} - o_j^{(t)}| \leq \delta \\ o_i^{(t)} & \text{if } |o_i^{(t)} - o_j^{(t)}| > \delta, \end{cases} \tag{2.4}$$

where $\mu \in (0, 1/2]$ is a parameter that models how much is conceded upon interaction, and analogous, simultaneous update occurs for $o_j^{(t+1)}$. In contrast, the HK

variant considers synchronous updates of all agents in the society, with each agent being simultaneously influenced by all neighbours that are within their tolerance region at each time step,

$$o_i^{(t+1)} = |I(i,t)|^{-1} \sum_{j \in I(i,t)} o_j^{(t)}, \tag{2.5}$$

where $I(i,t) = \left\{ j \;\middle|\; |o_i^{(t)} - o_j^{(t)}| < \delta \right\}$. Therefore, an important difference between both variants is that the DW is stochastic while the HK is deterministic.

Since their conception, many variations and additions to the models and how they affect the transition points in δ have been studied. For instance, the inclusion noise in the form of nodes relocating to anywhere in the space with (as a notion of 'free will') prevents the formation of clusters at low values of δ (Pineda et al., 2009), while making the outcomes of the model less dependent on initial conditions (Carro et al., 2013). When agents are placed in a network structure, they can arrive at consensus for lower values of δ, which can be explained by the role of bridges in the network (Schawe et al., 2021), while the presence of zealots or radical agents who never change their opinion can produce an effect of polarisation (Hegselmann and Krause, 2015). Another important extension is the extension of the model to multiple dimensions, for which it is necessary to define a distance metric, although not much differences are found between the Manhattan and Euclidean distances (Fortunato et al., 2005). In two–dimensional spaces, opinions clusters gather at square lattices (Lorenz, 2008).

Other recent variations of the bounded–confidence models attempt to provide more realistic accounts of how states of polarisation emerge in societies by means of repulsion mechanisms and exogenous events. For instance, Macy et al. (2021) include mechanisms of repulsion, party identity, and exogenous shocks that unite opinions in a multidimensional model, and show that partisanship can create tipping points at states of polarisation by which the populations cannot recover, even after the partisanship is removed. Condie and Condie (2021) also study the effect of exogenous shocks in a networked population where individuals also have a 'certainty' in their opinion, only being convinced by individuals with higher certainty, and who also display random fluctuation in their opinions. They show that when shocks are not present the population arrive at a wide consensus, but the introduction of shocks creates a state of polarisation. Based on the idea that people may have incentives to express opinions more polarised than their real ones in social media, Lim and Bentley (2022) show that opinion amplification can also act as a mechanism that leads bounded–confidence–like dynamics to a state of high polarisation. However, they provide two simple intervention strategies — limiting the number of amplified messages allowed and spreading balanced opinions — that can effectively counter the polarising effect.

2.3 External influence of opinions dynamics

The basic statistical physics approach of most models outlined above is the study of complex behaviours emerging from simple interactions of the many constituents of the system. While this approach already provides insight into interesting social phenomena, a key element widely prevalent in social information systems is still missing: *external influence*. 'External influence' can be understood here in two different ways. First, as an agent that is outside the system in that it does not follow the opinion dynamics that all other agents in the system follow, which can be thought as 'internal' or 'intrinsic'. Hence, this agent has the power to affect but *not be affected* by it (Masuda, 2015; Brooks and Porter, 2020) or at least a disproportionate imbalance in the influence power (Kacperski and Holyst, 2000). However, unlike zealot agents introduced above, the external influence typically has the power to affect all or most individuals of the social group (Carletti et al., 2006) —a form of mass media or government propaganda — or to strategically choose whom to influence (Masuda, 2015; Bimpikis et al., 2016) —e.g. public policies. Second, external influence can be understood as a process that interferes the intrinsic dynamics of the social group by means of a different mechanism. This can be modelled e.g. as an external field (Lynn and Lee, 2016; Crokidakis, 2012) or a controller that is able to affect aspects of the social system such as the placement of zealots (Vendeville et al., 2022; Yildiz et al., 2013) or altering the social ties. Below, we divide external influence by whether their attempts are finely targeted to the level of picking individuals or edges to affect or whether they are untargeted or targeted very coarsely and review the literature covering each of the two categories.

2.3.1 Untargeted external influence (or coarsely targeted)

Untargeted external influence is typically referred to as *mass media*, or *media sources*, as they mimic in many cases the role of radio, television, newspapers, and other forms of information channels that reach big parts of the social group. Research questions here concern how untargeted external influence may affect the opinion dynamics, resulting in consensus to a different opinion to what would be expected without the external influence (Laguna et al., 2013; Crokidakis, 2012), or sometimes hindering (or fostering) the appearance of a consensus altogether (Colaiori and Castellano, 2015; Peres and Fontanari, 2011).

From a modelling perspective, untargeted external influence is typically introduced as an immutable agent with which all agents in the social group interact, which may occur after every peer interaction (Sîrbu et al., 2013b; Quattrociocchi et al., 2011), at other rates (Carletti et al., 2006; Vaz Martins et al., 2010), instead of peer interactions with a certain probability (Laguna et al., 2013; Pineda and Buendía, 2015; Lu et al.,

2009), or be taken as an agent that has connections to (most) agents in the network and follows normal interaction rules of the opinion dynamics model (Bhat and Redner, 2020). Although conceptually different, the '*truth*' (when opinions are related to perceptions of some truth) can have an identical role to mass media in opinion dynamics models (Kurz and Rambau, 2011), as it is an immutable source to which most agents have partial access —typically through a form of 'sensing' or reasoning (Brede and Romero Moreno, 2022).

External influence may produce different outcomes depending on the opinion dynamics model, with major qualitative differences between binary (or categorical) and continuous opinions. Among binary opinion models, many of them have been extended with untargeted external influence, such as voter–model–like models (Colaiori and Castellano, 2015), the Sznajd model (Crokidakis, 2012), the Ising model (Tessone and Toral, 2009), the Naming Game (Lu et al., 2009), or the Axelrod model (Peres and Fontanari, 2011). The effect of external media varies depending on its modelling details and the opinion dynamics. In some cases, the external influence can bring the social opinion to a consensus (Colaiori and Castellano, 2015), which in some cases it may be systematically aligned with the media's opinion (Laguna et al., 2013). In the Axelrod model of cultural dynamics and contrary to basic intuition, it has been shown that an external medium with a fixed cultural profile fosters diversity in the population (Peres and Fontanari, 2011). Tessone and Toral (2009) study IM for an Ising model that includes fixed preferred opinions in agents, finding that having diversity in the distribution of the preferred opinions facilitates arriving at a consensus around the media's opinion. A different picture emerges when more than two or more sources of media compete in the spread of ideas, typically leading to a fragmentation of opinions in the population (Bhat and Redner, 2020).

Many models with continuous opinions also include the role of external influence, ranging from models with repulsive forces (Sîrbu et al., 2013b), Kinetic models (Boudin et al., 2010), or the bounded confidence model (Carletti et al., 2006; Vaz Martins et al., 2010; Kurz and Rambau, 2011; Pineda and Buendía, 2015; Brooks and Porter, 2020; Quattrociocchi et al., 2015). Due to the segmentation dynamics produced by bounded confidence or repulsive forces, the effect of the external media here typically depends on its intensity. When an external medium aggressively promotes an extremist view, it may succeed in capturing a small fraction of agents that initially were in its vicinity, but most agents will likely be unaffected by its influence (Sîrbu et al., 2013b), or even cluster further to the other side of the opinion spectrum (Carletti et al., 2006). This relates both to the intensity of the message and the frequency of its broadcasting (Carletti et al., 2006). Pineda and Buendía (2015) found that populations with heterogeneous bounded–confidence thresholds are more sensitive to external influence. The inclusion of additional media sources with opposing extreme views, however, may foster a population consensus at a moderate opinion (Sîrbu et al.,

2013a), and moderate media outlets with links to only sub-parts of the population may have more success at gathering followers close to their views than greedier ones (Brooks and Porter, 2020).

In the above works, the external control is passive, although its effect is studied under different scenarios. In contrast, Albi et al. (2015) explore the idea of how external fields in kinetic models can be optimally controlled to bring the population towards the desired opinion. Other optimising works look at external control in the form of strategic agents that can modify their opinion as the dynamics unfold. Their goal may be to bring opinions in the population to a desired state (Mirtabatabaei et al., 2014; Hegselmann and Krause, 2015; Wongkaew et al., 2015) or to minimise the time it takes to reach consensus (Kurz, 2015). A different means of control is tampering with the opinion dynamic itself to enforce (or prevent) consensus in the society. Examples of this approach are the control of the order of pairwise encounters of agents (instead of random, (Lorenz and Urbig, 2007)), or controlling how much agents are allowed to move when updating their opinions (Li et al., 2020).

In contrast to the above, not much work has been done around external controllers that react to the evolution of the distribution of opinions to instantly maximise their influence. Such modelling is realistic in that external agents may have very limited information about the dynamics of opinions, what other opponents may be doing, or may have bounded rationality. An account of this approach is given by Quattrociocchi et al. (2015), where several media outlets interact with citizens affecting their opinions. Further to the dynamics among citizens, the media opinions also evolve according to their own coupled dynamics, copying the most successful of their neighbours (the one that has more citizens within their BC interval) as they compete for followers. They find that such media competition leads to a polarisation of opinions in the society.

2.3.2 Finely targeted external influence

Finely targeted external influence refers here to an external agent that performs local modifications in the population in an attempt to indirectly influence the whole system. The modelling of external influence may vary, from the selection of '*seed*' agents that are automatically converted into zealots (Kuhlman et al., 2013; Yildiz et al., 2013; Arendt and Blaha, 2015; Liu et al., 2010; Gionis et al., 2013), to choosing targets that are linked to a zealot that represents the external influence (Masuda et al., 2010; Brede et al., 2018; Chakraborty et al., 2019; Lynn and Lee, 2016; Lu et al., 2009; Ghezelbash et al., 2019; Mai and Abed, 2019; Yi and Patterson, 2020), or modifying the network structure (Gaitonde et al., 2020a; Musco et al., 2018; Chitra and Musco, 2020; Dong et al., 2017). Targeting only a subset of the population is typically sought because it may lead to better results, or because a more resource–intensive (and more effective) form of communication is modelled and a limited budget needs to be considered.

Hence, with a targeted external influence, it is paramount to determine its optimal placing, a problem that is commonly termed *influence maximisation* (*IM*).

Studies on IM allow us to better understand different aspects of influencing attempts, such as under which conditions one can effectively introduce the desired opinion in a population (İkizler, 2019; Albanese et al., 2020), how influencing strategies interact when competition for influence takes place (Goyal et al., 2019; Hoferer et al., 2020), or what the effect of external influence is on reaching consensus in a population (Bhat and Redner, 2020). Influence maximisation has huge relevance for a variety of important applications that range from advertising (Domingos and Richardson, 2001), political and public information campaigns (Wilder et al., 2018a; Zhang et al., 2016), to questions on how to optimally encourage a developing economy (Alshamsi et al., 2018b).

Influence maximisation has already been studied for almost two decades, starting with the seminal works of Domingos and Richardson (2001), Richardson and Domingos (2002) and Kempe et al. (2003). Their work employed binary, progressive models —the independent cascade (IC) and linear threshold (LT)— which remained as the main preference on following studies in the computer science literature (Li et al., 2018). Kempe et al. (2003) initially defined influence maximisation in the context of progressive opinion models as finding the k-subset of initial active nodes (*seeds*) that achieves the largest expected number of activated nodes at the end of the progressive diffusion process. They proved that this problem is NP-hard for both the IC and LT models, and provided a greedy hill–climbing algorithm that approximates the optimal solution within a factor of $1 - 1/e$ ($\sim 63\%$) in polynomial time. The whole process requires evaluating the spread of influence for every node in the network with k repeats. However, since an exact calculation of the spread of influence is exponential in the network size[5], the task is only feasible for medium–size networks via approximated evaluations that simulate many instances of the diffusion process with Monte Carlo sampling. Therefore, existing work of influence maximisation in the computer science community has mainly been devoted to improving the efficiency of the greedy algorithm in both time and accuracy (for a review, see Li et al. (2018) and Peng et al. (2018)). Most works tackle the standard problem of a single influencer trying to bring their opinion to the network, although a smaller stream of work has focused on the game–theoretical scenario with two adversarial influencers (Budak et al., 2011; Tsai et al., 2012), providing best–response actions to the problem for different graph settings and under the possibility of uncertainty.

In non-progressive models, nodes can change their opinions back and forth and hence the diffusion process may not arrive at a static attractor. For this reason, the influence maximisation goal needs an additional temporal specification, namely achieving the highest influence spread at a specific time horizon (Brede et al., 2019a; Cai et al., 2021),

[5]It has been shown to be #P-hard (Chen et al., 2010a,b).

at the asymptotic equilibrium (if it exists, (Masuda, 2015)), or along the whole diffusion process (Kempe et al., 2003). Although in most cases the goal of the external influence is to bring the population to a desired opinion, some studies also consider alternative goals such as the promotion of consensus (Lu et al., 2009) or polarisation (Ghezelbash et al., 2019).

Regarding the opinion diffusion models, IM has been studied in many of them, including both binary/categorical and continuous models. Among binary models, IM has been studied for the Voter Model (Kuhlman et al., 2013), the AB model (Arendt and Blaha, 2015), the SIR and its variants (Kandhway and Kuri, 2014), the Ising model (Liu et al., 2010), or the Naming Game (Lu et al., 2009). For continuous models, IM has mainly been studied for the DeGroot model (Ghezelbash et al., 2019) and the Friedkin–Johnsen (Abebe et al., 2018) Similar to progressive models, IM is solved in many cases for non-progressive models via a greedy algorithm (Liu et al., 2010). Other studies approach solutions to IM via gradient ascent algorithms (Lynn and Lee, 2016; Mai and Abed, 2019), message–passing algorithms (Vassio et al., 2014), percolation theory (Saito et al., 2012) Shapley Values (Jain et al., 2021), topology–based heuristics (Montes et al., 2020) or other analytical approaches (Eshghi et al., 2020; Kandhway and Kuri, 2014). Additionally, in many cases, IM is considered against an opponent, which may be passive (Kuhlman et al., 2013) or actively maximising their influence, for which a game–theoretical approach is needed (Goyal and Manjunath, 2020; Dhamal et al., 2018).

Next, the literature related to IM for the Voter Model is reviewed in more detail, as it is closer to this work.

Influence maximisation in the voter model

The first studies performing influence maximisation on the Voter Model were performed by Even-Dar and Shapira (2011). In their problem definition, the external controller can influence the initial opinion of some individuals in the network (under some budget constraint) and attempts to maximise the expected number of nodes holding the desired opinion in the steady state. They provide an exact solution to the problem when the cost of setting the initial opinion of a node is uniform among the population, which is picking the nodes with the highest degree of the network. Masucci and Silva (2014) extended this work to a competitive setting with two external controllers favouring opposite opinions and Li et al. (2013) to signed networks, with positive and negative links. These works face some problems as the steady state of the Voter Model becomes independent of the initial conditions under some realistic modifications of it, such as the inclusion of noise (Brede et al., 2019b) or zealots (Masuda, 2015). Therefore, modifying initial opinions could only be useful when maximising opinions for a short time horizon instead of in the steady state.

Due to the important role that zealots play in the composition of opinions in equilibrium, some works conceive IM in the Voter Model as the optimal placement of zealots in the network (Kuhlman et al., 2013; Yildiz et al., 2013). A first approach was to consider which individuals should be transformed into zealots of the desired opinion to maximally spread it in the steady state (Kuhlman et al., 2013; Yildiz et al., 2013). Note that, in the Voter Model, the presence of zealots holding the same opinion guarantees its reaching the whole population. For this reason, these studies assume that zealots of the opposing opinion are randomly scattered in the social network. Comparable to the question of finding optimal seeds in the independent cascade model, this approach assumes that nodes can be readily transformed into zealots of the desired opinions without considering the costs and effort of conversion, which is too strong an assumption.

A different approach, introduced by Masuda (2015), takes inspiration from other works on network control (Porfiri and di Bernardo, 2008; Liu et al., 2011) as a way of bringing a network with a dynamics process happening between its nodes to the desired state by introducing an external influence in a subset of the nodes. In this line, Masuda treats zealots as external agents who exert influence via strategically placing control links to nodes in the social network. Unlike the above zealot approaches, in this framework the external influence needs to compete with peer influence at the entry points, so nodes targeted by the external zealot do not deterministically stay at their side. This approach is closer to the common scenarios of viral marketing and mass media, as users targeted by these techniques are not guaranteed to be convinced. Influence maximization is then conceptualised as identifying an optimal set of targets subject to a budget constraint and against a passive, opposing controller with random targets. Masuda (2015) has shown that, for undirected networks, optimal targets typically follow a degree ordering, starting with nodes of the largest degrees. Subsequent works have employed the same framework to show that noise or copying errors (Brede et al., 2019b) and resistance to attempts of control (Brede et al., 2018) can shift optimal targeting to low–degree nodes in certain parameter regimes. Similar effects can also be observed if optimization is not aimed at achieving maximum vote shares in the stationary state, but rather at a finite time horizon (Brede et al., 2019a). Then, when time horizons are short, control cannot always capture hub nodes in time and targeting lower degree nodes may become optimal. All these studies with external controllers use stochastic hill–climbing algorithms to find the set of nodes that maximise influence without providing a theoretical guarantee for the quality of the approximated solution.

All of the above studies that include zealots assume a discrete targeting of the controller; a node is either a zealot or not (Kuhlman et al., 2013; Yildiz et al., 2013), or an edge is either built to a node or not (Masuda, 2015; Brede et al., 2018, 2019b,a). However, controllers may want to split their budget unevenly between nodes to

achieve a higher impact. Furthermore, most works cited above hold an algorithmic approach to the influence maximisation problem: they focus on developing algorithms that reach optimal solutions with proven bounds on their approximation factors and complexity scaling. However, very little attention is placed on how the structure of solutions looks like, i.e. which nodes are preferentially targeted and under which conditions. For instance, there are well–known heuristics that make use of the structural properties of nodes and perform reasonably well for influence maximisation (Montes et al., 2020; Erkol et al., 2019). If optimal algorithms perform better than such heuristics, where does this extra advantage come from?

2.4 Agent–Based Modelling

Agent-Based Modelling (ABM) are a simulation modelling techniques in which individual entities with a set of rules of behaviour are let to interact. They are particularly useful for the study of complex systems that are dynamic and highly non-linear, resulting in chaotic or emergent phenomena. They are a theoretical exercise in that no empirical data is used, but they do generate data is later analysed, so they can be considered as a 'third way of doing science' (Axelrod (1997a), p. 3). Most of the opinion models above can be approached via ABM, as they depart from microscopic rules of agent behaviours that can be simulated. However, many of them are typically designed to be mathematically tractable and researchers typically mix analytical and numerical approaches in their study. However, if models grow in complexity, ABM becomes the most appropriate approach for their analysis.

As it is generally common in any modelling approach, models should not grow arbitrarily in complexity; an unmoderated increase of ingredients would difficult the inspection of the model and the understanding of the mechanisms that produce its behaviour. Furthermore, ABM simulations may be computationally costly, so a focus in parsimony is also desired to alleviate computational burden. Therefore, it is important to only keep the model ingredients that are thought to be relevant and contribute to the dynamics of the studied phenomenon, while removing irrelevant details. Further challenges in the implementation of ABM include risks that 'interesting' behaviour emerges from unnoticed errors in the computer code that implemented the model. Due to the complex nature of the models, these code errors may be hard detect, so rigorous programming practices should be employed. Likewise, a rigorous tracking of model changes, and code reusability are key to ensure the reproducibility of the experiments (Axelrod, 1997a). Last, there are risks of numerical stabilities that come from any simulation task, as computational representation of quantities have a finite limit in their precision and the biggest and smallest quantities that they can represent, and care must be put in detecting these situations (Hegselmann and Krause, 2015).

Earliest cases of ABM can be found in cellular automata — machines formed by a grid of cells that change state by simple rules according to the state of the neighbours — which were able to produce chaotic and complex behaviour. These models were taken and expanded in the field of population biology and developed into a sub-research field called *artificial life*, where biological and ecological process are typically studied by creating environment with resources and granting agents the abilities to interact with others and the environment and to reproduce (Langton, 1995). Paradigmatic examples of ABM applied to ecological phenomena include the study of swarm behaviour — such as flocking, schooling, or foraging — how cooperative behaviours can emerge from evolution (Smith, 1976; Axelrod and Hamilton, 1981), or how 'swarm intelligence' emerges from distributed behaviours in colonies of ants and bees. These were also naturally transferred to the design of *swarm robotics* which, inspired by biological populations, attempt to increase the robustness of a group of robots by having decentralised behaviours (Brede and Romero-Moreno, 2022).

In parallel of the emergence of the field of artificial life, ABM found a natural ground in the study of social systems, where it was applied to diverse phenomena, such as neighbourhood seggregation (Schelling, 1971), the emergence of norms (Axelrod, 1997a), or the collapse of civilisations (Roman et al., 2017). The first large–scale ABM was developed by Epstein and Axtell (1996), featuring complex societies of agents including very diverse phenomena such as migration, pollution, reproduction, combat, trade, diseases, and culture (Epstein and Axtell, 1996). Many other large–scale ABM representing societies at the level of cities or states followed with a wide variety of topics of study, from political violence (Cioffi-Revilla and Rouleau, 2010) to pandemic outbursts (Panovska-Griffiths et al., 2022).

Related to the topic of opinion dynamics and external control, a stream of ABM research has also focused on developing models in the context of political science, with political parties and mass media are viewed as types of agents that interact with the opinion dynamics of the agents as normal citizens (Sobkowicz, 2016; Moya et al., 2017). Particularly interesting are the cases around party competition, where the opinion space reflects a space of political stances and parties compete for votes. According to the classic theory of issue voting from Downs (1957), voters will select their chosen party depending on their distance to the party platforms in this political space. Consequently, parties — whose survival is strongly tied to their capability to secure votes — can act strategically to maximise their vote share, either by moving their platforms (typically at a cost in votes or in utility if moving away of the party's core values) or by affecting the political issues that are perceived as most important by voters when comparing their distances to the parties, i.e. by 'setting the agenda' (De Sio and Weber, 2014).

The Downsian model predicts that both parties should converge to the median voter's position (Downs, 1957), which is at odds the behaviour observed in many political

systems. Inspired the research using ABM in evolutionary game theory (Axelrod, 1997a), the application of ABM to party competition showed that more realistic outcomes of divergence in party platforms emerged if parties are under *bounded rationality*, i.e. they have limited information about voters' preferences or the opponent's strategies, having an 'adaptive behaviour' instead of being perfect optimisers (Laver and Schilperoord, 2007). Another landmark of the use of ABM in contrast to more classic techniques from economics is that it allows to study the dynamic aspects of adaptive learning instead of focusing on the equilibrium point of forward–looking strategic analysis (Laver and Sergenti, 2011; GARCÍA-DÍAZ et al., 2013)

The study of Kollman et al. (1992) was among the first in which parties move their platform to maximise votes. They implemented three adaptive strategies for parties to change their platform: i) *Random Adaptive Parties*, who are able to poll L points of small changes in party platform and move to the one that gives them the highest expected vote share, ii) *Climbing Adaptive Parties*, who perform L hill–climbing iterations of changes in party platform, and iii) *Genetive Adaptive Parties*, who perform $L/2$ iterations of a genetic algorithm — where L represents the length of the campaign previous to each election. They also implemented two classes of parties: *ambitious parties*, who only seek to maximise votes, and *ideological parties*, who prioritise maximising votes when they are not the majority, but also seek to remain closer to their ideology represented by their initial position. They showed that, although the performance of the three strategies did not differ much, ideological parties have much higher chance to remain incumbent than ambitious ones. Following Kollman's work, Miller and Stadler (1998) showed that hill–climbing party strategies asymptotically converge to the median voter, although times to reach there may be long and rich, chaotic–like transient dynamics can emerge before that point is reached. Later models explored additional party strategies (LAVER, 2005; Fowler and Laver, 2008) and extensions, such as the endogenous appearance of new parties (Laver and Schilperoord, 2007), the interplay of party competition and social turnout (Fowler and Smirnov, 2005), changes in the dimensionality of the political space (Garcia et al., 2012), or the formation of coalitions (Lehrer and Schumacher, 2018).

A different approach to party competition is by analysing how parties alter the perceived importance of specific issues (Budge and Farlie, 1983). This approach is related to the agenda–setting capabilities of parties, who are able to influence the topics that are discussed by influencing the mass media to promote aspects of the debate in which they hold favourable positions relative to the electorate (De Sio and Weber, 2014). Hence, the assumption is that, when computing the distance of a voter to a party, the different component dimensions of the political space are weighted by the perceived importance (or, political sciences terms, *saliency*), which is the one that parties want to affect (Amorós and Puy, 2010). Interestingly, Dragu and Fan (2016)

showed that this form of party competition led parties in the minority to support divisive topics within the electorate. Meyer and Wagner (2019) posited that the movement of party platforms is only an effect of aggregating changes in saliency in low–level issues into coarser political dimensions, unifying both perspectives. In contrast, other works have simultaneously considered movement of party platforms and changes in saliency to study the appearance of new parties (Tavits, 2008) and the success of niche parties (Meguid, 2005). Muis (2010) models both party competition in platform and media effects on voter perceptions, although their media distortion represents a lack of media coverage that introduces increases the perceived distance from voters instead of pursuing an agenda–setting effect. Crucially, some of these works have made efforts to calibrate model parameters with empirical data or used it for validation (LAVER, 2005; Muis, 2010; Sobkowicz, 2015; Moya et al., 2017; Muis and Scholte, 2013).

2.5 Optimisation methods

Optimisation is a branch of mathematics that involves obtaining the highest (or lowest) value of an objective quantity by choosing values of a set of variables. Optimisation occurs at many levels in nature — as emerging behaviours — such as in societies (trading), biology (evolution), and physics (energy) (Nocedal and Wright, 1999). From a human decision perspective, optimisation techniques allow to operate on mathematical models of the world, typically with the help of computers, and are a key element of engineering. An optimisation problem can be formally defined as[6]

$$\max_{\boldsymbol{x}} \; f(\boldsymbol{x}), \qquad s.t. \quad g_i(\boldsymbol{x}) = 0 \quad h_j(\boldsymbol{x}) \geq 0\,, \tag{2.6}$$

where $\boldsymbol{x}$ is the vector or variables, f denotes the objective function that links the variables to the objective function, and g_i and h_j a set of functions specifying constraints in the values of the varaibles that can be picked.

Approaches and methods to optimisation vary according to the nature of the variables (e.g. continuous or discrete, bounded or unbounded) and the type of function that relates them to the objective (e.g. implicit or explicit, linear or non-linear, known or unknown). Depending on the complexity of the problem and the method employed, the optimisation task may require computation above the resources available, maybe a solution is not even feasible, or there may not be guarantees that the solution obtained is the optimal one. Thus, some important properties that characterise optimisation methods are the following.

[6]Note that while the equation refers to max, all concepts apply to seeking the minimal value; it is enough to maximise the function $f' = -f$.

1. **Time complexity,** or how much time the optimisation method takes to obtain a solution. It typically depends on the characteristics of the problem, such as its size (in terms of the number of variables and other parameters). This dependencies are typically expressed with a *big O* notation, $O(X)$, which represents a computing time that scales linearly with the expression X in the worst–case scenarios (upper bound).

2. **Memory complexity,** or how much information the optimisation method needs to hold in memory simultaneously. Similarly to the time complexity, it can be expressed with a *big O* notation.

3. **Guarantees on error bounds** in the form of a limit in how far the solution obtained is from the optimal one or, in the case of discrete variables, a probability that the optimal solution was found.

4. **Robustness.** The quality of the solution should not vary considerably with the initial conditions (in iterative methods).

Therefore, a large body of research is devoted to develop and understand different optimisation method that are appropriate for varying contexts depending on resources avaiable and needs.

A major difference in the nature of optimisation problems and solutions stems from the space of the variables. If variables are restricted to discrete values (e.g. binary, categorical, integer), we talk about *discrete optimisation*, while *continuous optimisation* refers to optimisation with variables within the real numbers. We will cover each of these next.

2.5.1 Discrete optimisation

Discrete optimisation involves an objective function that depends on a set of discrete variables, i.e. binary, categorical, or integer numbers. Problems with binary or categorical variables without an order (sets) may lack any information about how to explore the space of solutions and exploring all possible combinations of variables may be the only guarantee that an optimal solution is found. Such problems are said to be *NP-hard*, and their time complexity scales as Z^N, where Z is the number of categories per variable and N is the number of variables. Hence, ensuring that an optimal solution is obtained is only feasible for small problems with low number of variables N.

If the space cannot be fully explored due to limitations of resources, it is not possible to know whether the solution obtained is the optimal one. By exploiting the structure of a problem or domain knowledge, some methods employ smart heuristics to discard

parts of the space that are unlikely to contain the optimal solution and therefore can arrive at it with higher chances. However, it is still not possible to know whether it has indeed been found. *Stochastic hill–climbing* is a different approach which, starting from a guess, makes random changes in the variables and only keeps those if they lead to higher values of the objective function. Similarly, *greedy hill–climbing* starts from an initial guess and optimises each variable once, one at a time, performing the optimisation in a single pass. Under some properties of the function f, greedy hill–climbing algorithms can guarantee that the solution obtained is at least $1 - 1/e$ ($\sim 63\%$) times smaller than the optimal one. *Genetic algorithms* are a sophisticated extension of these methods that draws inspiration from evolutionary processes to create a population of solutions that then follow an iterated two–stage process. First, single values in each solution may change with some probability (mutation) and combinations of variables values may be exchanged between a pair of solutions (crossover). At the second stage, solutions are replicated (offspring) according to their evaluation in the objective function f (fitness) and those with the lowest fitness get discarded. Due to the role of crossover, genetic algorithms are able to better arrive at solutions to which hill–climbing would not be able. However, this comes at the cost of higher computational cost and additional mechanisms to ensure diversity in the population of solutions.

Problems whose variables are integer numbers are of a very different nature, since there is a sense of proximity between the different values that a variable can obtain. The set of methods that are used for solving these are collectively called as *Integer programming*, which is used in many areas of business and industry including logistics, finance, or manufacturing.

2.5.2 Continuous optimisation

Continuous optimisation involves optimising an objective function that depends on a set of continuous variables, i.e. they are real numbers. This type of optimisation is typically easier to achieve than discrete optimisation if the function to optimise is *smooth*, i.e. if its first and second derivatives are defined in all points of the variable space. This smoothness allows to know the local behaviour of the function around any specific point, providing valuable information about in which direction its values are increasing or decreasing and the rate of change (gradient). Continuous optimisation methods generally start from an initial guess $\boldsymbol{x}_0$ and use the information of the gradient of the function ∇f to navigate the space in the direction of increasing values until a maximum is found, i.e. a point with higher than all points in its neighbourhood. However, this may not be enough, as there may be points with even higher values further apart — i.e. it may be a *local maximum*.

Often, it is not possible to know whether the maximum arrived at is a local or a global one; only a wider exploration of the variable space can provide some hope that the solution found is at least 'good enough'. However, there is a class of problems, called *convex problems* which guarantee that there is only a single maximum, that is therefore the global one (Nesterov, 2004). Convex problems are characterised by a convex function f and a convex space of variables $\boldsymbol{x}$, where a convex function implies

$$f(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \leq \alpha f(\boldsymbol{x}) + (1-\alpha)f(\boldsymbol{y}) \qquad \forall\alpha \in [0,1] \tag{2.7}$$

and a convex variable space $\mathbb{S}$ implies

$$\boldsymbol{x} \in \mathbb{S},\ \boldsymbol{y} \in \mathbb{S} \rightarrow \alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y} \in \mathbb{S}\,. \tag{2.8}$$

Continuous optimisation problems move through the space via discrete steps (*iterations*) and they mainly differ in how these steps are calculated. A common approach is to compute a direction of movement and then find an appropriate step size at each iteration, a family of methods known as *line search* (Nocedal and Wright, 1999). An appropriate direction of improvement can be that in which the function will increase the most, which is the maximum direction of the gradient, what gives these methods the common name of *gradient ascent*.

The choice of the step size can come in many flavours. While too small step sizes will unnecessarily increase the number of iterations to reach the solution, too big step sizes can also be inefficient — or even create instability of the method, preventing it to ever converging at a solution — as the information conveyed in the gradient may only hold locally (Nocedal and Wright, 1999). The magnitude of the gradient can provide useful information for adapting the step size too: a large gradient magnitude indicates higher confidence that the function is increasing in that direction, while the gradient magnitude will become vanishingly smaller as a maximum is approached. To prevent overshooting, the value of the function f at iteration $i+1$ can be compared to that of i; if it is lower, that means that the step size should be reduced and a new step from the previous point should be computed — a technique called *backtracking* (Nocedal and Wright, 1999). An extension of this method checks the value of f at several step sizes and keeps the best of them, or even performing a full-fledge optimisation (by taking several steps) along that direction. However, there is a clear trade–off between the time spent to compute a good step size and how much improvement is obtained by it and the balance would come from how expensive is to evaluate f and to compute ∇f (Nocedal and Wright, 1999).

A more sophisticated technique that simultaneously provides the direction and size of the steps is the *Newton direction*, which uses the second derivative $\nabla^2 f$ to provide a more efficient navigation of the space (Nocedal and Wright, 1999). Therefore, this technique is recommended in problems with complex surfaces of f. Additionally, as it

provides optimal step sizes, problems arising from selecting inappropriate sizes can be avoided. However, the computation of $\nabla^2 f$ is not always possible, or it may be too computationally costly to perform it at every step.

2.6 Summary

This chapter has covered different approaches to mathematical modelling of influence networks, opinion dynamics, and external influence —both untargeted and targeted. The literature around these topics is vast and varied, with a profusion of modelling variations and little consensus about which models are the most appropriate.

To start with, opinions can be modelled as discrete or continuous variables, leading to different dynamics and system behaviours. To cover both branches, one model with binary opinion variables and one model with continuous opinion variables are included in this thesis. Regarding models with discrete opinions, there is a first distinction in the literature between dynamics where opinions change in only one direction (progressive models) and those where opinions can flip back and forth between different states (non-progressive). While progressive models are popular in IM for the computer science field, non-progressive models are more appropriate for opinions that do not require commitment or for regarding longer time scales (Arendt and Blaha, 2015). A paradigmatic non-progressive model is the Voter Model, which has a long history of research and presents high levels of tractability (Redner, 2019). For this reason, and since it has also been empirically proven to match voting data (Fernández-Gracia et al., 2014; Braha and de Aguiar, 2017; Vendeville et al., 2021), it will be employed in this thesis for opinion dynamics with binary opinions. Regarding models with continuous opinions, the bounded–confidence model certainly enjoys the widest acceptance and research tradition, both in general opinion dynamics and in the study of external influence in particular (Noorazar et al., 2020; Sîrbu et al., 2017; Castellano et al., 2009a), so it will be the choice as a model with continuous opinions.

External influence has also been studied from different modelling approaches, ranging from the introduction of zealots within the network to representations of the media as external agents that link to specific nodes or as tampering with aspects of the system, such as the initial opinions or the network structure. Most works related to IM assume discrete allocations of influence from the optimiser and apply a greedy hill–climbing algorithm to arrive at an approximate solution. Despite being proposed in the seminal works of IM (Domingos and Richardson, 2001; Kempe et al., 2003), continuous allocations have barely been regarded. Moreover, most works only focus on developing fast and accurate algorithms for solving IM while ignoring the shape of the solutions' structure. To cover these gaps, this work introduces a continuous IM problem where external controllers build links to agents in the network. The aim is to

develop solutions to the continuous allocation problem, study the structure of such solutions, and compare them with the solutions obtained from the standard discrete allocation setting in the literature.

Furthermore, we analyse the role of external influence in a different setting where the controllers react adaptively to a population that evolves in a consensus–formation process. For this scenario, we find motivation in the context of political parties that affect the relative importance that different political dimensions have when interacting with others and going to the polls, a well–studied phenomenon in the field of political science (Amorós and Puy, 2010; Feld et al., 2014; Dragu and Fan, 2016; De Sio and Weber, 2014). However, having external controllers affecting the perceptions of dimensions in a multi-dimensional opinion space is novel and has never been approached from the opinion dynamics field. To that end, we study the role of external agents that compete by affecting the importance mix of opinion dimensions in the context of a consensus–formation process of a population following the bounded–confidence dynamics.

Chapter 3

Continuous Influence Maximisation for the Voter Dynamics: Optimal solutions and analysis of the structure of optimal allocations

Parts of this chapter have been published as a journal paper (Romero Moreno et al., 2021a), an article in a conference proceedings (Romero Moreno et al., 2020c), and an extended abstract (Romero Moreno et al., 2020b).

3.1 Introduction

As we have seen in Sect. 2.3.2, previous research on IM —both on progressive and non-progressive models— has mainly employed a discrete approach, i.e. identifying a set of K individuals whose control maximises the share of the desired opinion (Domingos and Richardson, 2001; Zhang et al., 2016; Yadav et al., 2017; Kuhlman et al., 2013; Brede et al., 2018; Dezső and Barabási, 2002; Chen et al., 2009; Budak et al., 2011; Even-Dar and Shapira, 2011; Arendt and Blaha, 2015; Kempe et al., 2003; Lynn and Lee, 2018; Masuda, 2015; Brede et al., 2018). However, the discretisation of influence targeting is artificially constraining if campaigners wish to allocate different amounts of resources to different groups, e.g. weakly targeting one group of nodes while focusing strong controlling power on another subset of targeted nodes. The latter strategy can be better captured by a continuous approach where the external controller can tailor the intensity of influence to individual target nodes (Eshghi et al., 2020). This approach gives a more suitable model to describe mixed modes of

influence campaigns, which typically combine generic messages directed to a broad public (through radio, television, or billboards) with intense, bespoke promotion to specific groups of the population.

Although the assumption of discrete targeting is prevalent in most works on influence maximisation, the seminal works from Richardson and Domingos (2002) and Kempe et al. (2003) included the possibility of continuous allocation. Due to the linearity of their continuous allocation version of the problem, Richardson and Domingos (2002) obtained a closed–formed solution to the maximisation process. Kempe et al. (2003) made their influence maximisation problem continuous by probabilistically determining the initial seeds depending on the amount of budget allocated to them. Although they proposed an approximate gradient hill–climbing algorithm, the evaluation of their goal function is NP-hard and they do not provide a method for approximate evaluations. Closer to our work, Lynn and Lee (2016) study influence maximisation with continuous allocations for the Ising model and derive a mean–field approximation to obtain optimal influence allocations for the steady–state distribution given a passive opposing field (equivalent to an opposing external controller). They prove that their goal function is concave and propose a gradient ascent algorithm that obtains an ϵ-approximation to the optimal solution in $O(1/\epsilon)$ iterations. To the best of our knowledge, the only instance of continuous allocations applied for influence maximisation in the voter model happens in the competitive influence maximisation problem of Masucci and Silva (2014), where two players can continuously allocate their influence and a node's initial opinion is sided with the player that invested the most on it (an instance of a Colonel Blotto game). This approach only applies to competitive games and, as mentioned before, is not robust as initial conditions become irrelevant in the long run in the presence of zealots or noise (Masuda, 2015). To bridge the gap of continuous influence allocations in the voting dynamics, we formalise the problem of influence maximisation with continuous allocations (in Sect. 3.2) and propose a gradient–ascent algorithm to obtain optimal solutions numerically.

Furthermore, a thorough analysing of the structure of the optimal solutions obtained is prominently lacking in the IM literature, where the focus generally is on improving algorithmic performance and providing approximation guarantees (Li et al., 2018; Kempe et al., 2003; Yadav et al., 2017; Lynn and Lee, 2018). A main reason for this gap is the high dimensionality and complexity of the structure of solutions, as they include thousands of variables that are typically interconnected in complex network structures. Therefore, the rest of the chapter is devoted to a detailed analysis of the structure of optimal solutions to continuous IM, via analytical results and numerical techniques.

As a first finding, we establish that optimal allocations tend to spread across large parts of the network (Sect. 3.4). Second, we present two simple heuristics that can be derived from optimal solutions to continuous IM: a direct response to the nodes

targeted by the opponent, *shadowing* (Sect. 3.5), and an indirect response that focuses on the neighbours of the nodes targeted by the opponent, *shielding* (Sect. 3.6), which may be understood as first– and second–order approximations[1] to optimal solutions, respectively. By understanding these two types of heuristics, we can provide intuition on what are the key ingredients of an optimal solution to our continuous IM problem. Third, we explore to what extent optimal solutions in the continuous and discrete regimes can be explained by hub targeting, and how this popular heuristic is related to shadowing and shielding (Sect. 3.7). Fourth, we extend the analysis to the conventional discrete IM and assess differences between the solutions obtained in the two regimes (Sect. 3.8). Our results demonstrate that whereas hub preferences are very important in the discrete setting, shadowing and shielding heuristics are the predominant factors explaining maximal influence in the continuous case. Finally, we analyse the game–theoretical scenario where both external controllers actively adapt their strategy and provide an empirical understanding of the Nash equilibrium (Sect. 3.9), finding that a unique equilibrium lies on equal targeting of all nodes in the network. We further quantify the degree to which an opponent strategy can be exploited by an active controller and find that deviations from optimal allocations are more detrimental the lower the degree of the node at which they occur.

3.2 Formalisation of the continuous IM problem

The standard version of the voter model assumes a graph $G = \{V, E\}$ where agents are identified with nodes: $v_i \in V, i = 1, \ldots, N$, and edges are influence links between them: $e_{ij} = \{v_i, v_j\} \in E$. We assume that the network is positively weighted, undirected, and has no self–loops, with $W = (w_{ij})$ corresponding to its weighted adjacency matrix. Each agent holds one of two possible opinion states $o_i \in \{A, B\}$, with $i = 1, ..., N$. Agents update their opinions subject to the voter dynamics: at every iteration, a node is chosen at random and copies the opinion of one neighbour who is chosen with probability proportional to the weight of their common link.

Further to the standard model, external influence is introduced via zealots as external controllers (Masuda, 2015; Brede et al., 2018; Chakraborty et al., 2019; Cai et al., 2021). We assume that there are two external controllers holding different opinions. These controllers exert influence on the network via unidirectional links with weights w_{ai} (w_{bi}) that are also taken into account in the updating dynamics of opinions —i.e. they count as neighbours when performing copying in the opinion dynamics. In the

[1] Note that we refer to 'order' here as the distance of node targeting from nodes targeted by the opponent. In this sense, shadowing can be considered a first–order heuristic because it is related to allocations given to nodes targeted by the opponent. On the other hand, shielding constitutes a second–order heuristic because it considers nodes in the direct environment of nodes targeted by the opponent. Further–order shielding is also possible, although not explored in depth here due to its higher complexity and lower effect.

continuous regime, controllers can freely decide which nodes they target, with positive continuous strengths $w_{ai}, w_{bi} \in \mathbb{R}^+$ and subject to a budget constraint, $\mathcal{B}_{a(b)} \geq \sum_i w_{ai}(w_{bi})$. The discrete regime is a sub-case of the continuous regime that only allows two possible weight values, $w_{ai}, w_{bi} \in \{0, g\}$, where g is a fixed *gain*. The number K of nodes thus targeted is related to the budget via $K = \lfloor \mathcal{B}/g \rfloor$.

As presented in Sec. 2.3.2, the behaviour of the opinion dynamics under the voter model can be studied by considering the evolution of probabilities $x_i(t) \in [0, 1]$ that a node holds opinion $o_i = A$ at time t (Mobilia et al., 2007; Even-Dar and Shapira, 2011; Masuda, 2015). From this perspective, the dynamics of a node can be described by the rate equations

$$\frac{dx_i}{dt} = (1 - x_i)\frac{\sum_j w_{ij}x_j + w_{ai}}{d_i + w_{ai} + w_{bi}} - x_i\frac{\sum_j w_{ij}(1 - x_j) + w_{bi}}{d_i + w_{ai} + w_{bi}}, \tag{3.1}$$

where d_i is the weighted degree of a node, $d_i = \sum_j w_{ij}$.

Here, we focus on maximising influence in the steady–state of the distribution. As Masuda (2015) has shown, this system has exactly one equilibrium point $\boldsymbol{x}^*$ that is an attracting state, meaning that the expected asymptotic distribution of opinions over the network does not depend on the initial opinion distribution. It can be found via the linear relationship

$$(L + W_a + W_b)\,\boldsymbol{x}^* = \boldsymbol{w}_a\,, \tag{3.2}$$

where L is the weighted Laplacian of the network, W_a (W_b) is a diagonal matrix whose diagonal entries correspond to w_{ai} (w_{bi}), and bold symbols are column vectors. The total vote share of nodes holding opinion A at the equilibrium is obtained from

$$X = \frac{1}{N}\sum_i x_i^* = \frac{1}{N}\mathbf{1}^T\boldsymbol{x}^* = \frac{1}{N}\mathbf{1}^T(L + W_a + W_b)^{-1}W_a\mathbf{1}\,. \tag{3.3}$$

To simplify notation, we will use x_i and $\boldsymbol{x}$ to refer to the single fixed point corresponding to x_i^* and $\boldsymbol{x}^*$ in the following.

For the IM problem, without loss of generality, we assume that the controller favouring A is active, i.e. it seeks to find the best way to distribute its link weights, $\boldsymbol{w}_a$, with an aim to maximise the vote share of its opinion, X. We generally assume that the opposing controller, favouring B, is passive, with links $\boldsymbol{w}_b$ that are fixed and known by the active controller, and that controllers have full information about the network structure. The IM problem with continuous allocations and for the steady–state can then be formulated as

$$\max_{W_a}\; X, \qquad s.t.\quad X = \frac{1}{N}\mathbf{1}^T(L + W_a + W_b)^{-1}W_a\mathbf{1}, \quad \sum_i w_{ai} \leq \mathcal{B}_a, \;\; w_{ai} \geq 0\,. \tag{3.4}$$

A first step towards solving the continuous IM optimisation problem is finding the critical points where gradients of the goal function are equal to zero, corresponding to

$$\nabla_{\boldsymbol{w}_a} X = \frac{1}{N}\Big[\mathbf{1}^T \cdot (L + W_a + W_b)^{-1}[\, I - \mathrm{diag}(x_i)]\Big]^T = \mathbf{0}\,, \tag{3.5}$$

where $\mathrm{diag}(x_i)$ is a diagonal matrix whose diagonal elements correspond to the probability vector at the steady–state $\boldsymbol{x} = (L + W_a + W_b)^{-1} W_a \mathbf{1}$. Note that all elements in $\nabla_{\boldsymbol{w}_a} X$ are positive, since $(L + W_a + W_b)$ is an M-matrix (Slawski and Hein, 2015), which are a subset of the inverse-positive matrices (Fujimoto and Ranade, 2004).

Since $\nabla_{\boldsymbol{w}_a} X = \mathbf{0}$ is generally nonlinear in $\boldsymbol{w}_a$ —and analytically intractable for the general case— we propose two different approaches to finding a solution: a numerical solution via gradient ascent for the general case (Sec. 3.3.1), and analytically studying a heterogeneous mean–field approximation (Sec. 3.3.2), which simplifies analytical calculations under some assumptions.

3.3 Methods and experimental settings

In this section, we present our methodological approach for finding solutions to the IM problem. We propose two different approaches for the continuous allocation version of IM: a **numerical** solution, via a gradient–ascent algorithm (Sec. 3.3.1), and an **analytical** solution, obtained via a mean–field approximation that simplifies the model by lumping nodes with a same degree into classes of nodes showing similar behaviour (Sec. 3.3.2). Last, Sect. 3.3.3 presents the experimental settings of the remaining sections of the chapter.

3.3.1 Gradient ascent algorithm for continuous IM in the voter model

The continuous modelling of IM brings a methodological benefit. Solving IM in discrete control framing is only feasible via combinatorial optimisation, whose exact solution scales exponentially in time with the number of targets. On the contrary, the assumption of continuous strengths allows for better mathematical treatment of the problem, since the problem becomes differentiable and local search techniques can be applied to finding an optimum. Moreover, due to the concavity of our IM problem in the strategy space (for a proof, see Appendix A.9), local–search techniques are guaranteed to arrive at the global optimum (Nesterov, 2004). Here, we employ gradient ascent as a local–search method, which has previously been used in IM by Lynn and Lee (2016). This technique is ensured to reach an ϵ-approximation to the exact solution in $O(1/\epsilon)$ iterations of the algorithm (Nesterov, 2004).

Our version of the algorithm is shown in Algorithm 1. Starting from a uniform allocation distribution, each iteration performs a step in the allocation space in the direction of the gradient, moving to allocations with increasing vote shares for the controller. Note that we have not included the budget and positivity constraint in the gradient calculation, so after every iteration, the allocation distribution needs to be projected back to the regular N-simplex that fulfils both constraints. The projection to the N-simplex is done following the algorithm developed by Chen and Ye (2011).

Algorithm 1: ϵ-approximation to optimal allocation of influence via gradient ascent

input : $\mathcal{B}_a, L, W_b, \mu, \epsilon$
output: approximation for $\boldsymbol{w}_a^*$ at the global maximum, X^*

1 $w_{ai}^{t=0} = \mathcal{B}_a/N; \quad X^{t=0} = 0;$
2 **repeat**
3 $\quad \nabla_{\boldsymbol{w}_a} X^{t-1} = \frac{1}{N}\left[\mathbf{1}^T (L + W_a^{t-1} + W_b)^{-1}[\, I - \mathrm{diag}(x_i^{t-1})]\right]^T$ (3.5);
4 $\quad \boldsymbol{w}_a' = \boldsymbol{w}_a^{t-1} + \mu \nabla_{\boldsymbol{w}_a} X^{t-1}$;
5 $\quad \boldsymbol{w}_a^t =$ Projection of $\boldsymbol{w}_a'$ onto the N-simplex constraint;
6 $\quad X^t = \frac{1}{N}\mathbf{1}^T (L + W_a^t + W_b)^{-1} W_a^t \mathbf{1}$ (3.3);
7 $\quad$ Backtracking;
8 **until** $X^t - X^{t-1} < \epsilon$;

The parameter μ (line 4 of Algorithm 1) regulates the step size. Big steps will generally speed up the process but may endanger its convergence. To prevent divergence, we also employ *backtracking* after each iteration; i.e. the new solution is rejected and the step size halved if the vote share is lower in the new iteration than in the previous one. Note that every step requires a $N \times N$ matrix inversion, so the time complexity to reach an ϵ-approximate solution scales with N and ϵ as $O(N^3/\epsilon)$.

A test of the quality of the algorithm can be found in Appendix A.2, where we compare its solutions to analytically obtained ones.

3.3.2 Heterogeneous mean–field approximation

As shown in (3.5), general analytical solutions to the continuous IM problem are hard to obtain. To cope with this, much of the previous literature has focused on developing optimisation algorithms (Medya et al., 2019; Nayak et al., 2019) or finding heuristics that are close to optimal (Abebe et al., 2018; Alshamsi et al., 2018b). Amongst such heuristics, degree–based heuristics have played a prominent role (Dezső and Barabási, 2002; Banerjee et al., 2013). We take the latter as inspiration and develop an approximation that reduces the number of degrees of freedom in the system by grouping nodes with the same degree. Note that this approximation is only effective for unweighted networks.

The heterogeneous mean–field (HMF) approximation assumes that, independent of the details of individual neighbourhoods, every node is coupled to a mean–field where the probability of having a node as a neighbour is proportional to their degree (Moretti et al., 2012; Carro et al., 2016; Hu and Zhu, 2017; Brede et al., 2019b). This approximation is not always valid but tends to perform well for not overly sparse networks in the absence of particular degree correlations. The HMF approximation assumes random connections between nodes and ignores higher–order correlations that might be present in real–world networks.

Under this approximation, the expected opinion state of a neighbour $\langle x \rangle$ is presumed to be the same for every node in the network and equal to

$$\langle x \rangle = \frac{1}{N} \sum_i \frac{d_i}{\langle d \rangle} x_i \,, \tag{3.6}$$

where $\langle d \rangle = 1/N \sum_i d_i$. Note that high–degree nodes are more likely to be encountered as neighbours and hence products by $d_i / \langle d \rangle$ affect the x_i. We thus obtain probabilities of adopting A in the steady–state as

$$x_i = \frac{d_i \langle x \rangle + w_{ai}}{d_i + w_{ai} + w_{bi}}. \tag{3.7}$$

But Eq (3.7) can be re-inserted into Eq (3.6), arriving at a self–consistency equation and leading to the closed–form expression of the vote share as

$$X^{\mathrm{HMF}} = \frac{1}{N} \left(\sum_i \frac{d_i}{d_i + w_{ai} + w_{bi}} \right) \left(\sum_i \frac{d_i w_{ai}}{d_i + w_{ai} + w_{bi}} \right) \left(\sum_i \frac{d_i (w_{ai} + w_{bi})}{d_i + w_{ai} + w_{bi}} \right)^{-1} + \\ + \frac{1}{N} \sum_i \frac{w_{ai}}{d_i + w_{ai} + w_{bi}} \,. \tag{3.8}$$

Finding the optimal allocation in this setting is equivalent to finding $\boldsymbol{w}_a^*$ such that $\nabla_{\boldsymbol{w}_a^*} X^{\mathrm{HMF}} = 0$, leading to a system of high-order polynomial equations which is still hard to solve. Further conditions or assumptions will be presented for its solution later, namely a Taylor expansion around limits of low and high allocation strengths as compared to node degree (Sec. 3.5). An assessment of the accuracy of the HMF approximation on random graphs with bimodal degree distribution can be found in Appendix A.1).

3.3.3 Experimental settings

For illustration, most experiments shown in this chapter are performed on a real–world email interaction network (Guimerà et al., 2003), which provides a sample of a network topology related to social networks. This heterogeneous network is

unweighted and undirected, with a unique component of size $N = 1133$, mean degree $\langle d \rangle = 9.62$, and degree assortativity $\delta = 0.078$. Its degree distribution can be seen in Appendix. A.3. However, the results shown here also apply to more general classes of complex networks, as it is demonstrated in further experiments on other network topologies as shown in Appendix A.8.

Regarding the strategies of the passive opponent, we assume them to be non-strategic in most experiments, focuses her influence evenly among a number K of nodes in the network chosen at random, although some experiments also explore other strategies, such as other placings of the K nodes based on network properties, continuous allocations of strength spreading the whole network, or an opponent who also reacts strategically to the active controller. Influence budgets are fixed such as their sum equals to $\mathcal{B}_a + \mathcal{B}_b = N\langle d \rangle / 30$, entailing that, if both budgets are equal and controllers spread their allocations evenly through the network, the controllers' opinion would be copied around once every sixty times on average. This forces controllers to be strategic in the placing of their influence.

We obtain optimal solutions in the continuous regime numerically via the gradient ascent algorithm, employing $\mu = 50$ as the step size and $\epsilon = 10^{-10}$ as the convergence threshold, with all solutions converging in less than 10^5 iterations.

Optimal solutions for the discrete regime are obtained numerically via stochastic hill–climbing — a local–search technique which has been previously used for solving discrete IM in the voter model (Masuda, 2015; Brede et al., 2018, 2019b). The stochastic hill–climbing algorithm departs from a random subset of K nodes and iteratively swaps one of them with a random node in the network. The modified subset is preserved after each iteration if it performs better than the previous subset, i.e. if it leads to a higher X (the vote share to the active controller); otherwise, the change is reversed. This algorithm is typically run for a given number of iterations or until there has been a specific number of iterations without improvement. We perform stochastic hill–climbing for $20N$ iterations or until no change is made in the last $10N$ iterations, whichever occurs last.

To facilitate the analysis and description of the optimal allocation profile of numerical solutions, we have divided allocations into disjoint groups by targeting strength w_{ai} with respect to the mean allocation $\langle \boldsymbol{w}_a \rangle$. We have thus distinguished four disjoint groups $\mathbf{G} = \{\mathbf{H}, \mathbf{A}, \mathbf{V}, \mathbf{Z}\}$ of nodes with high allocations $\mathbf{H} = \{i \mid w_{ai} \geq 8\langle \boldsymbol{w}_a \rangle\}$, above–average allocations $\mathbf{A} = \{i \mid 8\langle \boldsymbol{w}_a \rangle > w_{ai} \geq \langle \boldsymbol{w}_a \rangle\}$, below–average allocations $\mathbf{V} = \{i \mid \langle \boldsymbol{w}_a \rangle > w_{ai} > 0\}$, and zero allocations $\mathbf{Z} = \{i \mid w_{ai} = 0\}$. We have chosen this particular partitioning to characterise the four distinct groups that can be observed in the optimal allocation profile from Fig. 3.1**b**, which will be used as a reference; this grouping is mainly employed for illustrative purposes for providing intuition, results are later extended to general cases. To explore the effect of the local topology around

nodes (related to the shielding behaviour), we define a second set of disjoint groups $\mathbf{S} = \{\mathbf{T}_b, \mathbf{N}_b, \mathbf{R}\}$ with nodes that are targeted by the passive controller ($\mathbf{T}_b = \{i \mid w_{bi} = g\}$), direct neighbours of nodes targeted by the passive controller ($\mathbf{N}_b = \{i \notin \mathbf{T}_b \mid (\exists j \in \mathbf{T}_b \mid w_{ij} > 0)\}$) and the remaining nodes ($\mathbf{R} = \{i \notin (\mathbf{T}_b \cup \mathbf{N}_b)\}$. We use the entropy of the allocation distribution as a characterisation of the spread of optimal allocations, $H(\boldsymbol{w}_a) = -\log(N)\ \sum_i (w_{ai}/\mathcal{B}_a) \log(w_{ai}/\mathcal{B}_a)$. Note that this entropy is normalised to the interval $[0, 1]$, which is achieved by the pre-factor $\log(N)$.

In Sect. 3.8, we employ heuristics inspired by the optimal solutions found in previous sections and to compare their effectiveness among the continuous and discrete regimes. We use four heuristics related to the discrete regime and four related to the continuous regime, whose implementation details we include below.

- *Random.* This is a baseline heuristic for the discrete regime, where K random nodes are targeted.
- *Degree–based (discrete).* This heuristic targets the K nodes that have either the highest or the lowest degree in the network, whichever results in the highest vote share.
- *Shadowing (discrete).* This heuristic targets exactly the same nodes as the passive controller.
- *Shielding (discrete).* Discrete shielding targets K nodes randomly selected from $\mathbf{N}_b$, although giving preference to nodes that have multiple neighbours in $\mathbf{T}_b$. If $K > |\mathbf{N}_b|$, the remaining targets are directed to random nodes in $\mathbf{R}$.
- *Uniform.* This is the baseline strategy for the continuous regime, targeting all nodes in the network with equal strength $w_{ai} = \mathcal{B}_a/N$.
- *Degree–based (continuous).* For this heuristic, allocation strengths are linearly proportional to node degree, $w_{ai} \propto k\, d_i$. The proportionality constant k is found numerically via binary search to maximise vote shares. Note that the *uniform* heuristic above is a sub-case of this heuristic with $k = 0$, so this heuristic will always perform equal or better than *uniform*.
- *Shadowing (continuous).* In this heuristic, there are only two possible allocation strengths, one given to nodes in $\mathbf{T}_b$ and one given to the remaining nodes. Their values are determined numerically via binary search to maximise the vote share. Note that *uniform* is a sub-case of this heuristic in which both intensities are equal, so this heuristic will always perform equal or better than *uniform*.
- *Shadowing plus shielding (continuous).* This last heuristic is a combination of both shadowing and shielding heuristics, where nodes are targeted with one of three possible intensities associated to each group in $\mathbf{S} = \{\mathbf{T}_b, \mathbf{N}_b, \mathbf{R}\}$. The values of

these intensities are determined numerically via binary search to maximise the vote share. Note that *shadowing* is a sub-case of this heuristic in which the allocations to nodes in $\mathbf{N}_b$ and $\mathbf{R}$ are equal, so this heuristic will always perform equal or better than *shadowing* (and, consequently, also better than *uniform*).

Note that the proposed heuristics require different amounts of information for their implementation. Some require the details of the network structure and/or knowledge about the opponent's strategy. Others have free parameters and their implementation requires an exploration of the parameter space, with computations of expected vote shares for evaluating each value of the parameters. Table 3.1 summarises what information is required for each of the heuristics along with the nested structure.

TABLE 3.1: **Information required by each heuristic (three middle columns) and nested structure (right column).**

	Network structure	**Opponent's strategy**	**Parameter Tuning**	**Nested heuristic**
Discrete heuristics				
random	-	-	-	-
degree–based	X	-	X	-
shadowing	-	X	-	-
shielding	X	X	-	
Continuous heuristics				
uniform	-	-	-	-
degree–based	X	-	X	*uniform*
shadowing	-	X	X	*uniform*
shadowing plus shielding	X	X	X	*shadowing*

The code for running all experiments and produce all figures in the chapter can be found at `https://git.soton.ac.uk/grm1g17/continuousim`.

3.4 Initial exploration of the structure of optimal allocations in the continuous IM

We first qualitatively explore the shape of optimal allocations in the continuous regime. The continuous regime allows for richer allocation decisions as controllers can reach all nodes in the network and modulate the strength of the allocation given to each node. If linked to practical examples, this would correspond to campaign managers using diverse modes of campaigning, potentially mixing a soft campaign directed at all individuals in a population with stronger campaigning directed at specific groups. Here, we address the question, if campaign managers would benefit from using such diverse modes of campaigning. For an initial exploration, we set the passive controller to target K random nodes in the network with equal strength. In

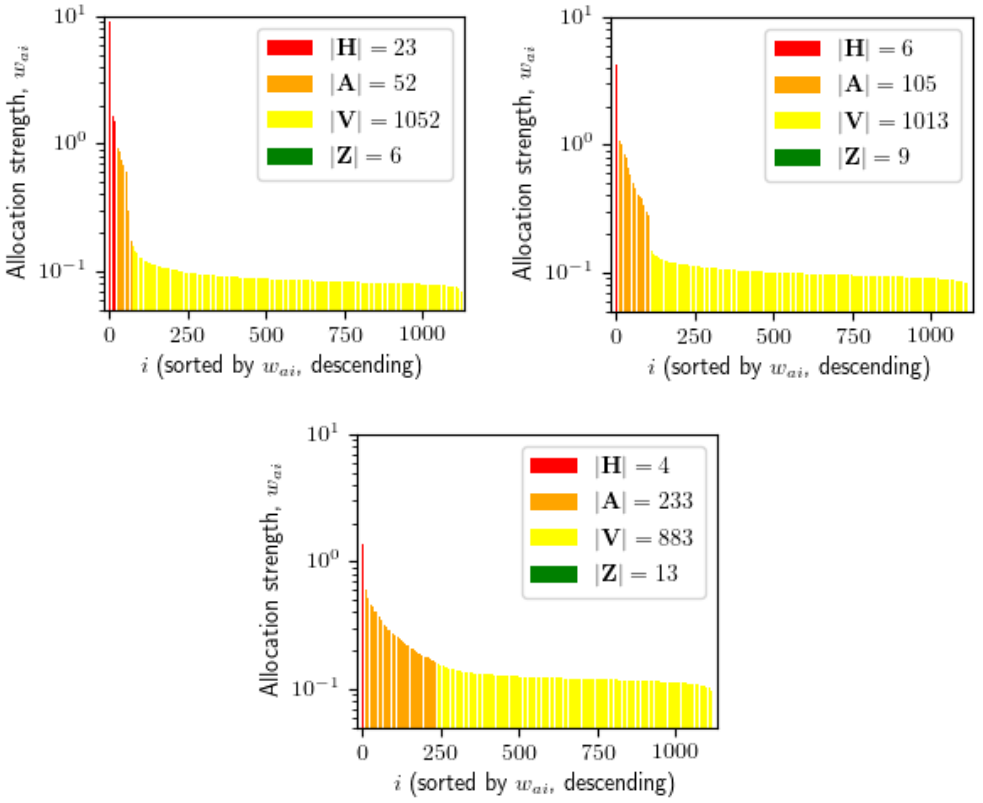


FIGURE 3.1: Distribution of optimal influence allocations sorted in descending order when the passive controller discretely targets **a** $K = 8$, **b** $K = 16$, and **c** $K = 32$ randomly chosen nodes. Both controllers hold the same budget, $\mathcal{B}_a = \mathcal{B}_b = N\langle d\rangle/60$. Allocations are coloured by the allocation group in $\mathbf{G} = \{\mathbf{H}, \mathbf{A}, \mathbf{V}, \mathbf{Z}\}$ they belong to.

Fig. 3.1, we present optimal control allocations sorted in descending order and obtained via gradient ascent. Results shown in the three panels of the figure have been obtained from numerical experiments where the passive controller targets $K = 8$, $K = 16$, and $K = 32$ nodes, respectively, and both controllers hold equal allocation budgets.

If a controller targeted K nodes in a discrete fashion, allocations to those K targeted nodes would belong to the high–allocation group $\mathbf{H}$ (if $K < N/8$) and allocations to the remaining nodes would belong to the zero–allocation group $\mathbf{Z}$. In contrast, we consider that a controller makes use of the continuous flexibility if allocations are spread across the network (with few nodes belonging to $\mathbf{Z}$) and cover a wide range of strengths. Fig. 3.1**b** illustrates that even though there is a focus of heavy targeting on some selected nodes ($|\mathbf{H}| + |\mathbf{A}| = 111$) and some nodes remain untargeted ($|\mathbf{Z}| = 9$), the allocation profile of the remaining nodes ($|\mathbf{V}| = 1013$) is mostly flat, with little variation of allocations around their mean, $std(\boldsymbol{w}_{a,i\in\mathbf{V}})/\langle \boldsymbol{w}_a\rangle_{i\in\mathbf{V}} = 0.11$. From these results, we see that the active controller is making use of the continuous flexibility, as allocations are widely spread across the network —only $|\mathbf{Z}| = 7$ nodes did not receive allocation at all— and across strengths —ranging from $w_{ai} = 10^{-1}$ to $w_{ai} = 5$. Looking

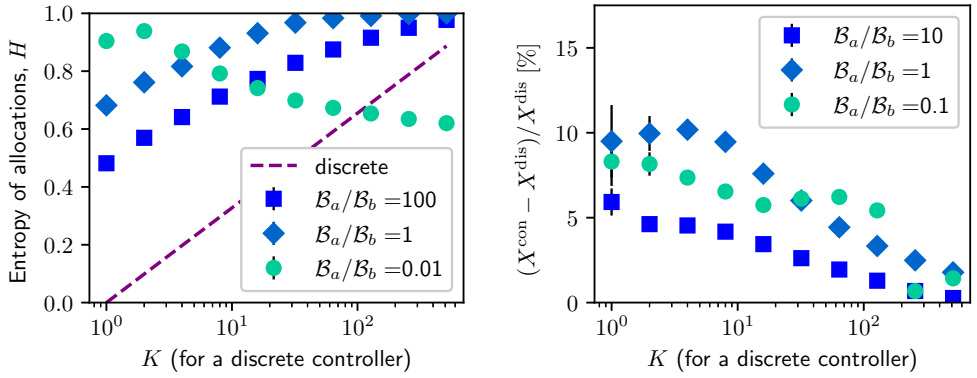


FIGURE 3.2: **a** Entropy of optimal allocation distributions in response to passive controllers with discrete allocations that are more or less widely spread (as measured by the number of nodes K that they target) and different budget ratios. The entropy of discrete allocations of the passive controller is also depicted (*dashed line*). **b** Percentage enhancement of control by optimal continuous allocations over optimal discrete allocations for varying K and budget ratios. Error bars represent standard errors over 15 experiment samples.

at other values of K (left and right panels from the figure), we note that the allocation distribution profile flattens as K increases, with a sharper peak when the opponent targets $K = 8$ nodes (Fig. 3.1–**a**), and a much smoother slope with $K = 32$ (Fig. 3.1–**c**). However, an almost–constant spread of allocations across all nodes in the networks remains present in all cases.

To extend the analysis of the allocation profile to other scenarios, we use the entropy of the allocation distribution, $H(\boldsymbol{w}_a)$, which characterises the spread of optimal allocations. Controllers that make use of the targeting flexibility of the continuous regime will tend to spread their allocations across the network, resulting in higher entropies. Fig. 3.2**a** shows that, almost independent of K and budget ratios, distributions of continuous allocations have considerably larger entropy than those of their discrete counterparts. We further note that the degree of spread is affected by the sparseness of the passive controller: the larger K and thus the more widely spread the strategy of the passive controller is, the more wide–spread optimal allocations are as well. In contrast, relative budgets ratios do not have a major effect on the entropy of the allocation distribution.

We next evaluate possible improvements obtainable by campaigning with continuous influence allocations, as measured by the relative improvement in vote shares gained over optimal discrete allocations, $(X^{cont} - X^{disc})/X^{disc}$. In Fig. 3.2**b**, we see that relative improvements are generally in the range of 5–10% against passive controllers with $K < 100$, while relative improvements drop below 5% against wide–spread passive controllers with $K > 100$. Note that, although the improvements can be small, they may be of significant importance in winner–takes–all scenarios, such as political

elections or referendums. Extensions to other network topologies of experiments from Figs 3.1**b** and 3.2**c** can be seen in Fig. A.11 of Appendix A.8, showing similar results.

To summarise, we have seen that spreading allocations over the whole network brings a bigger benefit to an optimal controller than when concentrating them on a few nodes. We also see that spreading allocations widely is optimal regardless of how many nodes are targeted by the opponent or the budget ratio between controllers. However, we also note that in optimal configurations some nodes are targeted much more strongly than others, with a particular focus on fairly few selected nodes $|\mathbf{H}| + |\mathbf{A}| = 111$, which comprise around 10% of the network. We next explore how these nodes are chosen, i.e. if they can be related to any of the shadowing, shielding, or hub preference heuristics.

3.5 First–order optimal responses: Shadowing (or avoidance)

Here, we build on a numerical observation of a rule to achieve optimal control in Brede et al. (2019a) that allocates control toward nodes also influenced by the passive controller (shadowing) or allocates control avoiding nodes targeted by the passive controller. Below, we shall show that shadowing is generally effective when the optimizer has a larger budget than the passive agent, while avoidance is more effective otherwise.

In this subsection, we aim to develop analytical intuition behind the shadowing behaviour for general network topologies. Due to the analytical intractability of the IM problem on complex networks, we use the heterogeneous mean–field (HMF) approximation to the opinion dynamics presented in Sec. 3.3.2. In this approximation, we assume that, independent of the details of individual neighbourhoods, every node is coupled to a mean–field. The vote share obtained by the approximation (Eq. 3.8) is non-linear in w_{ai}, so finding analytical expressions for optima is challenging. Therefore, to get analytical solutions, we focus here on the limiting cases of small and large external allocations with respect to the connectivity of the network. We thus perform Taylor series expansions in the limits of $(w_{ai} + w_{bi}) \ll d_i$ and $(w_{ai} + w_{bi}) \gg d_i$, which allow obtaining closed–form solutions for optimal strategies.

We start by analysing the low allocations limit, $(w_{ai} + w_{bi})/d_i \ll 1$, departing from the HMF approximation of the total vote share at the steady–state:

$$X^{\mathrm{HMF}} = \frac{1}{N}\left(\sum_i \frac{d_i}{d_i + w_{ai} + w_{bi}}\right)\left(\sum_i \frac{d_i w_{ai}}{d_i + w_{ai} + w_{bi}}\right)\left(\sum_i \frac{d_i(w_{ai} + w_{bi})}{d_i + w_{ai} + w_{bi}}\right)^{-1} + \\ + \frac{1}{N}\sum_i \frac{w_{ai}}{d_i + w_{ai} + w_{bi}} \,. \quad (3.9)$$

The optimisation of X^{HMF} —including the budget constraint $\sum_i w_{ai} \leq \mathcal{B}_a$ as a Lagrangian— can be written as

$$\max_{\boldsymbol{w}_a} \tilde{X} = \max_{\boldsymbol{w}_a} \left[X^{\mathrm{HMF}} + \frac{\lambda}{N} \left(\mathcal{B}_a - \sum_i w_{ai} \right) \right] , \tag{3.10}$$

where λ is the Lagrange multiplier. As computing $\nabla_{\boldsymbol{w}_a} \tilde{X} = \mathbf{0}$ leads to a system of polynomial equations for which no explicit solution can be computed, we derive a first–order approximation to X^{HMF} in the limit of $\alpha_i = (w_{ai} + w_{bi})/d_i \to 0$. For easiness of calculations, we can rewrite X^{HMF} as

$$X^{\mathrm{HMF}} = \frac{1}{N} \left(\sum_i \frac{1}{1+\alpha_i} \right) \left(\sum_i \frac{w_{ai}}{1+\alpha_i} \right) \left(\sum_i \frac{w_{ai}+w_{bi}}{1+\alpha_i} \right)^{-1} + \frac{1}{N} \sum_i \frac{\alpha_i}{1+\alpha_i} \frac{w_{ai}}{w_{ai}+w_{bi}} ,$$

and then apply the Taylor approximation to it:

$$X^{\mathrm{HMF}} = X^{\mathrm{HMF}}\Big|_{\alpha_i=0} + \sum_i \alpha_i \left. \frac{\partial X^{\mathrm{HM}}}{\partial \alpha_i} \right|_{\alpha_i=0} + \dots ,$$

$$X^{\mathrm{HMF}} = \frac{\mathcal{B}_a}{\mathcal{B}_a+\mathcal{B}_b} + \frac{1}{N} \sum_i \frac{w_{ai}+w_{bi}}{d_i} \left(-\frac{\mathcal{B}_a}{\mathcal{B}_a+\mathcal{B}_b} + N \frac{-\mathcal{B}_b w_{ai} + \mathcal{B}_a w_{bi}}{(\mathcal{B}_a+\mathcal{B}_b)^2} + \frac{w_{ai}}{w_{ai}+w_{bi}} \right) + \dots$$

By restricting the approximation to its first order and solving $\partial \tilde{X}/\partial w_{ai} = 0$, we arrive at optimal allocations in the low–allocations limit, w^L_{ai}, as

$$w^L_{ai} = \frac{1}{2} \left(\frac{\mathcal{B}_a}{\mathcal{B}_b} - 1 \right) w_{bi} + \frac{\mathcal{B}_a+\mathcal{B}_b}{2N} - \frac{(\mathcal{B}_a+\mathcal{B}_b)^2}{2N\mathcal{B}_b} d_i \lambda . \tag{3.11}$$

We can find the value of the Lagrange multiplier λ by inserting w^L_{ai} in the budget constraint $\mathcal{B}_a = \sum_i w_{ai}$, leading to the final expression for w^L_{ai} as

$$w^L_{ai} = \frac{\mathcal{B}_a+\mathcal{B}_b}{2N} + \frac{1}{2} \left(\frac{\mathcal{B}_a}{\mathcal{B}_b} - 1 \right) \left[w_{bi} + \frac{d_i}{\sum_{j:w^L_{aj}>0} d_j} \left(\mathcal{B}_b - \sum_{j:w^L_{aj}>0} w_{bj} \right) \right] , \tag{3.12}$$

where the sums over $j : w^L_{a_j} > 0$ are an effect of the positivity constraint $w_{ai} \geq 0$. Note that

$$\sum_{i:w_{ai}>0} w_{bi} = \mathcal{B}_b \iff \lambda = 0 \iff (\mathcal{B}_b - \mathcal{B}_a)\, w_{bi} \geq \mathcal{B}_b\, (\mathcal{B}_a+\mathcal{B}_b)/N \;\forall i ,$$

implying that the right term in the square brackets evaluates to zero when $(\mathcal{B}_b - \mathcal{B}_a)\, w_{bi} \geq \mathcal{B}_b\, (\mathcal{B}_a+\mathcal{B}_b)/N$. Otherwise, the solution needs to account for the $w^L_{ai} = 0$ that are breaking the positivity constraint. These can be found by iteratively computing the sums over $\forall j$, finding all values $w_{ai} < 0$ that break the constraint, setting them to zero, and readjusting the remaining values.

From Eq (3.12), we can observe that optimal allocations on the low–allocation limit w_{ai}^L depend linearly on the allocations of the opponent w_{bi} and node degree d_i, while both terms multiplied by a common coefficient whose sign depends on the ratio of budgets: $c = (\mathcal{B}_a/\mathcal{B}_b - 1)$. If $\mathcal{B}_a > \mathcal{B}_b$, i.e. if the active controller is in advantage of resources, the coefficient c is positive and the active controller allocates more influence to the nodes that are strongly targeted by the opponent, which corresponds to the shadowing strategy as discussed above. On the contrary, if $\mathcal{B}_a < \mathcal{B}_b$, the coefficient c is negative and the active controller allocates more influence to the nodes that receive the least allocation by the opponent, corresponding to avoidance behaviour. Note that, when the positivity constraints are not active ($w_{ai}^L > 0,\ \forall i$), there is no dependence of optimal allocations w_{ai}^L on node degree d_i, as the right term within square brackets is cancelled. In such cases, optimal allocations only depend on the allocations of the opponent w_{bi} and the budgets $\mathcal{B}_a$ and $\mathcal{B}_b$. Since allocations from the opponent cannot be negative either, the positivity constraints of the active controller may only activate when $\mathcal{B}_a < \mathcal{B}_b$. If they become active, optimal allocations also depend linearly on node degree d_i with a negative coefficient c, giving more allocations to nodes with low degrees. Hence, there is no scenario in which this analytical solution favours high–degree nodes.

We proceed similarly in the opposite limit of large allocations with respect to node degree, $d_i/(w_{ai} + w_{bi}) \ll 1$. We depart from the optimisation problem in (3.10) and derive a zeroth–order approximation to X^{HMF} in the limit of $\alpha_i = d_i/(w_{ai} + w_{bi}) \to 0$. X^{HMF} can be rewritten as

$$X^{\mathrm{HMF}} = \frac{1}{N}\left(\sum_i \frac{\alpha_i}{1+\alpha_i}\right)\left(\sum_i \frac{d_i}{1+\alpha_i}\frac{w_{ai}}{w_{ai}+w_{bi}}\right)\left(\sum_i \frac{d_i}{1+\alpha_i}\right)^{-1} + \\ + \frac{1}{N}\sum_i \frac{1}{1+\alpha_i}\frac{w_{ai}}{w_{ai}+w_{bi}}\,,$$

and the Taylor expansion evaluated as

$$X^{\mathrm{HMF}} = X^{\mathrm{HMF}}\Big|_{\alpha_i=0} + \dots\,, \qquad X^{\mathrm{HMF}} = \frac{1}{N}\sum_i \frac{w_{ai}}{w_{ai}+w_{bi}} + \dots\,,$$

Solving $\partial\tilde{X}/\partial w_{ai} = 0$ leads now to optimal allocations in the high–allocations regime w_{ai}^H as

$$w_{ai}^H = \sqrt{\frac{w_{bi}}{-\lambda}} - w_{bi}\,. \tag{3.13}$$

We can again find the value of the Lagrange multiplier λ by reintroducing (3.13) into the constraint $\mathcal{B}_a = \sum_i w_{ai}$ and write the final expression for w_{ai}^H as

$$w_{ai}^H = \left(\mathcal{B}_a + \sum_{j:w_{aj}^H>0} w_{bj}\right)\frac{\sqrt{w_{bi}}}{\sum_{j:w_{aj}^H>0}\sqrt{w_{bj}}} - w_{bi}\,, \tag{3.14}$$

where the sums over $j : w^L_{a_j} > 0$ are again arising from the need to fulfil the positivity constraint $w_{ai} \geq 0$. Which w_{ai} are breaking the positivity constraint can also be found by iteratively setting all values that break the constraint to zero and readjusting the remaining values, as with optimal solutions in the low–allocation limit.

In the high–allocation limit, optimal allocations are dependent on allocations of the opponent $\boldsymbol{w}_b$ —in a non-linear fashion— and the available budget $\mathcal{B}_a$. This solution does not show any dependence on node degree d_i.

To verify the validity of the assumptions involved in the above approximations in both low– and high–allocations limits, we compare the resulting analytical optimal allocations from Eqs. (3.12) and (3.14) to those obtained via numerical experiments. To cover a wide range of values in w_{bi}, we set the passive controller to target nodes with strengths w_{bi} randomly drawn from a uniform distribution, and then rescaled to meet the budget constraint. In Fig. 3.3, we examine the dependence of optimal responses w_{ai} on allocations of the passive controller w_{bi} for scenarios in which allocations generally are smaller (left column), similar (middle column), or larger (right column) than node degree, and the active controller being in budget superiority (top row), or inferiority (bottom row). In the figure, results corresponding to the HMF approximation are given as lines, while numerical results are given as clouds of points.

We observe that, when the active controller is in budget superiority (top row), the agreement between numerical and analytical results is very good, particularly in the limiting cases (cf. panels **a** and **c**). When allocations are similar in magnitude to node degree (Fig. 3.3**b**), numerical results are bounded by the analytical results for both limiting cases.

In contrast, when the active controller is in budget inferiority (bottom row), positivity constraints are active, so the term that depends on d_i activates in the low–allocations limit. This is particularly visible for low allocations (Fig. 3.3**d**), where allocations follow different trajectories depending on their degree. This effect gradually merges to the *dashed curve* related to the solution in the high–allocation limit (c.f. Fig. 3.3**e**), which matches quite well numerical results when allocations increase in magnitude (c.f Fig. 3.3).

3.6 Second–order optimal responses: Shielding (or anti–shielding)

Above, we have studied first–order responses to opposing strategies, which account for the most part of the distribution of optimal allocations. This is particularly true for continuous (or well–spread) allocation distributions, where we have seen that analytical solutions made by Taylor approximations fit very well numerical results.

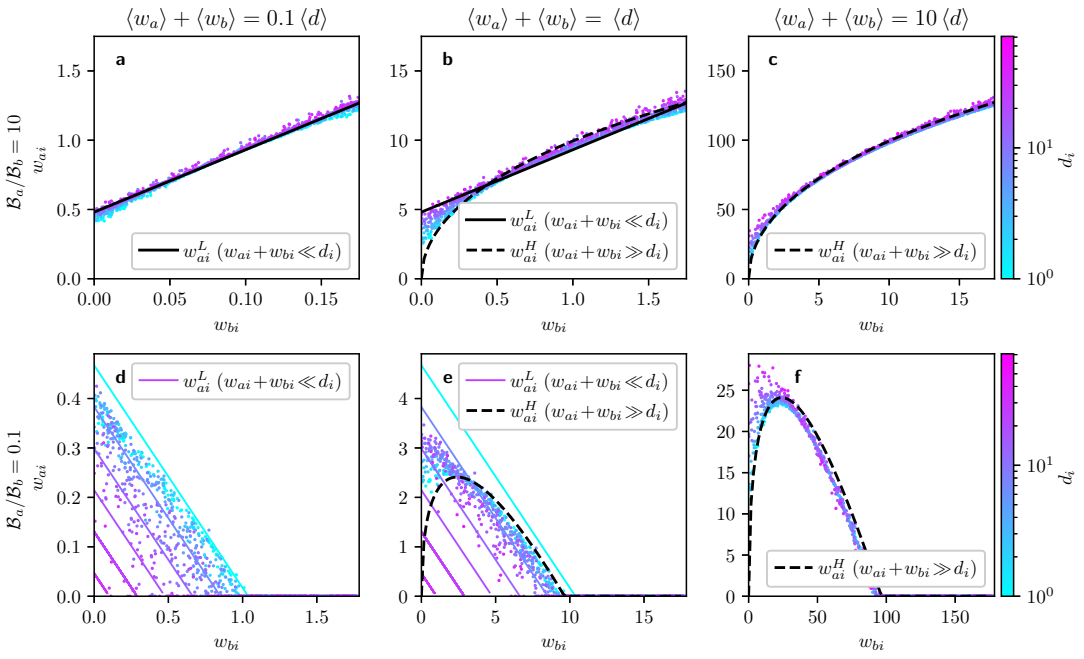


FIGURE 3.3: Comparison of analytical and numerical results related to shadowing for optimal allocations on an email–interaction network. Each panel shows the dependence of optimal allocations w_{ai} on the allocations of a passive controller w_{bi} resolved by classes of nodes of different degrees, indicated by colour. Columns show cases where the value of external allocations are generally lower (*left*), equal (*middle*), or higher (*right*) than node degree, as shown on the top of each panel. Rows show scenarios where the active controller is in budget superiority with $\mathcal{B}_a/\mathcal{B}_b = 10$ (*top row*), or in budget disadvantage with $\mathcal{B}_a/\mathcal{B}_b = 0.1$ (*bottom row*). The passive controller targets nodes with flexible strength, with weights randomly drawn from a uniform distribution. Numerical results are given by a cloud of points coloured by node degree and analytical results are given by *curves* that correspond to Eqs. (3.12) (*solid lines*) or (3.14) (*dashed curves*). Lines corresponding to w_{ai}^L are also coloured by node degree whenever the dependence on node degree activates, while showing only one line every six degrees (i.e. for degrees $d_i = 1, 7, 13, \ldots$).

However, such intuition is not satisfactory enough for opponents that target very few nodes in the network in a discrete manner. In such cases, we need to complement explanations of optimal allocations with second–order responses (i.e. shielding), for which we distinguish node classes depending on the distance to the targets of the opponent.

The idea behind shielding is that nodes that are direct neighbours of the nodes targeted by the opponent should receive higher allocations than other nodes in the network. The presence of shielding is then indicated by a strong dependence of allocations to nodes on the distance to nodes targeted by the opponent. To check the presence of this effect, we analyse here optimal allocations obtained numerically via gradient ascent and explore variations on the number of nodes targeted by the opponent (K). Figure 3.4 shows three cases in which the opponent targets **a** $K = 4$, **b** $K = 32$, and **c** $K = 256$ random nodes in a discrete manner. We can see there that

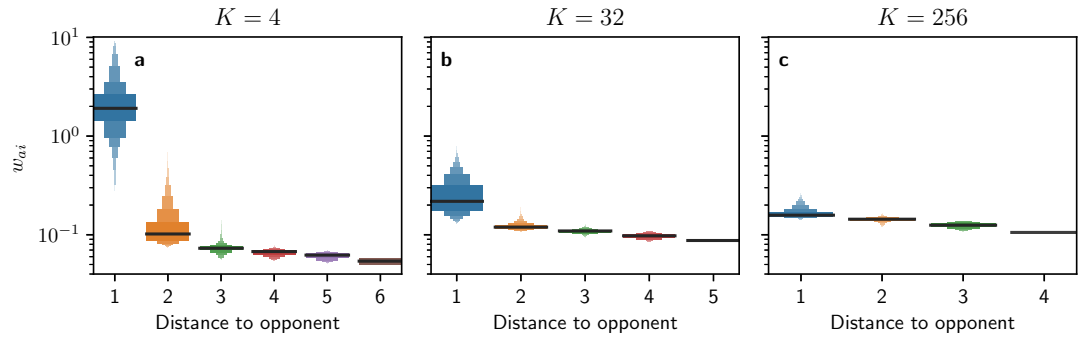


FIGURE 3.4: On the y axis, numerical optimal control allocations of the active controller in the form of a letter–value plot (Hofmann et al., 2017). On the x-axis, shortest distance to a node targeted by the passive controller who targets **a** $K = 4$, **b** $K = 32$, or **c** $K = 256$ random nodes in the network in a discrete fashion. Different levels of a boxplot represent the median, quartiles, octiles, and so on, while containing points from 15 repeats of the experiment.

shielding is strongly present when the passive opponent targets a few nodes in the network (c.f. panel **a**), but its effect decreases with K (c.f. panel **c**, where differences in allocations barely vary with the distance). This effect can also be seen in other network topologies, as shown in Appendix A.4.

If we go back to Fig. 3.1 in Sec. 3.4 and the allocation groups **G** there defined, we may wonder how shadowing and shielding can contribute to their explanation. We use a set of disjoint groups related to shielding $\mathbf{S} = \{\mathbf{T}_b, \mathbf{N}_b, \mathbf{R}\}$ with nodes that are targeted by the passive controller ($\mathbf{T}_b$), direct neighbours of nodes targeted by the passive controller ($\mathbf{N}_b$) and remaining nodes (**R**) and explore the overlap between the shielding groups in **S** and the allocation groups in **G** for the case where the passive controller targets $K = 16$ random nodes in the network (Fig. 3.5**a**). We observe that nodes that receive very high (**H**) or zero (**Z**) allocation are mainly nodes targeted by the passive controller ($\mathbf{T}_b$), the nodes from **A** that receive above–average allocations are mainly the neighbours of the nodes targeted by the passive controller ($\mathbf{N}_b$), and nodes receiving below–average allocations (**V**) correspond to those in **R**. This result demonstrates that differences in allocations given to nodes in the network can be largely explained by the combination of shadowing and shielding.

In Figs 3.5**b** and 3.5**c**, we extend the analysis beyond allocation groups in **G** and to other values of K by exploring average allocations $\langle w_a \rangle$ given to the shielding groups in **S**. When in budget equality (Fig. 3.5**b**), we observe that average allocations given to nodes in $\mathbf{T}_b$ and $\mathbf{N}_b$ are much larger than those given to nodes in **R** for $K < 50$, although this difference diminishes for higher K. These results show that both shielding and shadowing have a significant presence for low K. However, it is worth recalling from Fig. 3.5**a** that allocations given to nodes in $\mathbf{T}_b$ receive either high or zero allocations, so its distribution is bimodal and not well captured by the mean value. Extensions to other network topologies of experiments from Fig. 3.5**b** can be seen in

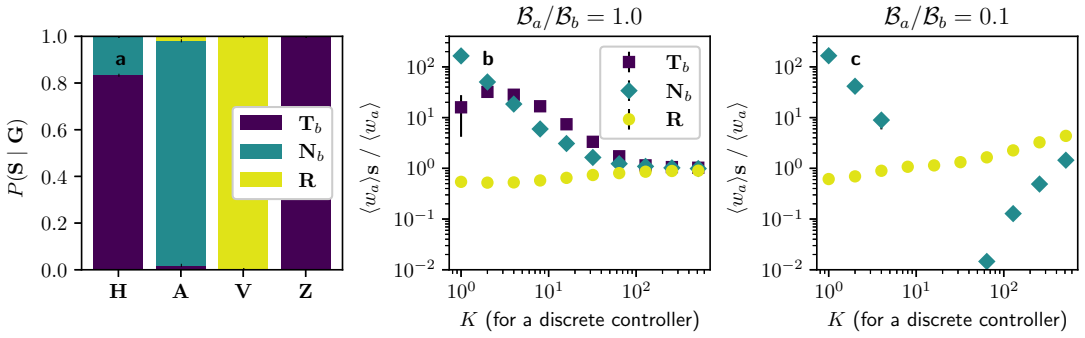


FIGURE 3.5: Analysis of shadowing and shielding heuristics on optimal influence allocations in the continuous regime for an email–interaction network. **a** Overlap between the allocation groups in **G** and shielding groups in **S** for a passive controller targeting $K = 16$ nodes in a discrete fashion. **b, c** Average allocations given to each group in **S** with respect to the total average allocation $\langle w_{ai} \rangle$ for different number of nodes K targeted by the passive controller and for **b** equal budgets or **c** an active controller in budget disadvantage. Error bars in **b, c** represent standard errors calculated from 15 repeats of the experiment. In **c**, points for $\mathbf{N}_b$ in the interval $K \in [5, 35]$ and for $\mathbf{T}_b$ in the whole range do not appear due to their value being equal to zero (and the scale being logarithmic).

Fig. A.12 of Appendix A.8, pointing to analogous results. Fig. 3.5**c** shows the same scenarios as in Fig. 3.5**b** but with the active controller being in budget disadvantage. Here, nodes in $\mathbf{T}_b$ receive zero allocation, corresponding to the avoidance behaviour discussed above. Interestingly, in this scenario nodes in $\mathbf{N}_b$ also receive less allocation $\langle w_a \rangle$ than those in **R** for $K > 5$, so avoidance also appears as a second–order strategy when in budget disadvantage. So direct (first–order) avoidance of the nodes targeted by the opponent is complemented by a second–order avoidance (*anti-shielding*) provided that the budget disadvantage is large enough. The case of budget advantage is not shown here, as it is very similar to that of budget equality, only with larger allocations given to nodes in $\mathbf{T}_b$.

3.7 Hub preferences and dependence on node degree

The third heuristic that we investigate is hub–targeting or, more generally, a degree–preference heuristic. We have seen that shadowing and shielding can explain much of optimal influence strategies in the continuous regime. However, we note that the effects of shielding and node degree are related, as nodes with high degrees are more likely to be neighbours of nodes targeted by the opponent and hence should also receive higher allocations due to shielding. Can a degree–preference heuristic explain optimal allocations more accurately than shielding? Or is the degree–preference just a heuristic that approximates the effect of shielding in many situations?

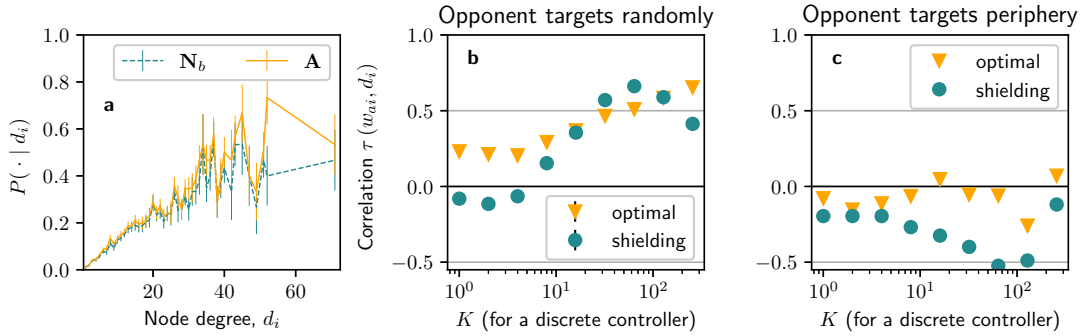


FIGURE 3.6: Correlations between optimal influence allocations and node degree on an email–interaction network. **a** Probability of optimal allocations belonging to $\mathbf{A}$ given node degree d_i (*solid*) and probability of nodes belonging to $\mathbf{N}_b$ given node degree d_i (*dashed*), with a passive controller targeting $K = 16$ random nodes in a discrete fashion. **b, c** Kendall rank correlation coefficients τ between allocation strength w_{ai} and node degree d_i for different K when the active controller targets nodes optimally (*triangles*) or following a shielding strategy (*circles*). In **a** and **b** the passive controller targets random nodes in the network. In **c**, the passive controller targets the nodes whose neighbours have the lowest possible degree. Error bars represent standard errors over **a** 60 or **b** 15 instances of the experiments. Correlations in **c** correspond to a single instance of the experiment, as the strategy of the passive controller is deterministic.

To test the relation between shielding and a degree–preference strategy, we look at conditional probabilities of nodes to belong to certain groups depending on their degree d_i. The first conditional probability, $P(i \in \mathbf{A} \mid d_i)$, relates to nodes that receive above–average optimal allocation and tests whether there is any dependence of allocations on node degree. The second conditional probability, $P(i \in \mathbf{N}_b \mid d_i)$, explores the link between shielding and node degree. These probabilities are shown in Fig. 3.6**a** for experiments run on the email–interaction network. It can be seen that both probability profiles are very similar, so it is very likely that one of these two heuristics is an artefact of the other one and thus can be superseded.

We generalise the relation of influence allocations w_{ai} and node degree d_i by measuring Kendall rank correlation coefficients τ (Kendall, 1957) between both quantities. A strong preference towards nodes with higher degree would then be indicated by correlations close to $\tau = 1$. Fig. 3.6**b** shows correlations with node degree by optimal allocations (*squares*), and by allocations from a strategy purely based on shielding (*circles*) —i.e. equally splitting all resources among nodes in $\mathbf{N}_b$. We observe that both the optimal and shielding strategies show considerable correlations with node degree for $K > 30$, where correlations are generally above $\tau > 0.5$. The lack of correlations with node degree for low K and the similarity in correlations between both strategies suggest that a degree–preference heuristic is only effective when shielding is correlated to node degree. To further illustrate this point, we perform another set of experiments in which we ensure that shielding and node degree are not positively correlated. We achieve this by having the passive controller only targeting

nodes whose neighbours have the smallest possible degree in the network, thus effectively displacing the targets of shielding to nodes with low degrees. Indeed, as shown in Fig. 3.6**c**, the correlations of optimal allocations with node degree vanish in these scenarios, with values below $|\tau| < 0.25$, while correlations of the shielding strategy remain negative with values down to $\tau \approx -0.5$. Extensions to other network topologies of experiments from Fig. 3.6**b** can be seen in Fig. A.12 in Appendix A.8, showing similar results. Also, a plot directly comparing node degrees d_i and optimal allocations w_{ai} can be found in Appendix A.6.

In conclusion, a preference towards hubs is appreciable in optimal allocations in the continuous regime, as high–degree nodes have higher probabilities of receiving higher allocations than other nodes. However, we argue that these correlations with node degree are a side effect of the shielding heuristic since nodes with a high degree also have higher chances of being neighbours of nodes targeted by the opponent. When we investigate scenarios with low K or with the passive controller targeting peripheral nodes, we observe the correlations of optimal allocations with node degree vanish, while being similar to that obtained by a shielding strategy. Therefore, we conclude that the correlations with node degree seen here are an artefact of the shielding strategy; optimal allocations can be explained well by the shielding heuristics without the need for additional assumptions about degree preference. This point is further developed in the next section.

3.8 Comparison of heuristics and optimal strategies in the continuous and discrete regimes

In the previous sections, we have obtained optimal allocation strategies via numerical methods and have analysed the structure of these solutions, while comparing them with some reference strategies. From these analyses, we have evaluated the importance of the three proposed explanatory factors (degree preference, shadowing, and shielding) in the continuous IM by studying their presence in optimal allocations. In this section, we extend the study to the discrete regime and compare how the three proposed heuristics perform when the active controller is constrained to discrete allocations. While targeting hub nodes or nodes with high centrality are common heuristics that have been shown to perform well in the discrete allocation regime (Dezső and Barabási, 2002; Chen et al., 2009; Budak et al., 2011; Even-Dar and Shapira, 2011; Masuda, 2015; Arendt and Blaha, 2015; Erkol et al., 2019; Montes et al., 2020)[2], we have shown in Sect. 3.7 that the preference towards hub nodes may be a side effect of a

[2]Again, while taking into account the full topology has been shown to perform better than hub heuristics also in the discrete regime (Kempe et al., 2003; Yadav et al., 2017; Lynn and Lee, 2018), previous studies mainly provide algorithmic approximations to optimal solutions without examining details of the topology of resultant optimal influence allocations.

shielding behaviour. By applying the framework with the three proposed heuristic to the discrete regime, we aim to explore whether the reported preference towards hub nodes may also be linked to a shielding behaviour. An extension to the study of the role of these three heuristics in the discrete regime can be found in Appendix A.7.

In this subsection, we employ an inverse approach for studying the structure of allocations: we build a set of heuristics based on the three factors and measure the gap in vote share ΔX relative to optimal results. The smaller the gap ΔX of a heuristic to the numerically determined optimum, the stronger its contribution to optimal allocations. We employ the following heuristics, whose implementation can be found in Sect. 3.3, four heuristics related to the discrete regime: *Random* (targets random nodes), *Degree–based (discrete)* (targets nodes with highest or lowest degree), *Shadowing (discrete)* (targets the same nodes as the opponent), and *Shielding (discrete)* (targets nodes in $\mathbf{N}_b$), and four related to the continuous regime: *Uniform* (targets all nodes with equal strength), *Degree–based (continuous)* (allocation strengths are linearly proportional to node degree), *Shadowing (continuous)* (different allocation strengths given to nodes in $\mathbf{T}_b$ and than the remaining nodes), and *Shadowing plus shielding (continuous)* (different allocation strenght given to each group in $\mathbf{S} = \{\mathbf{T}_b, \mathbf{N}_b, \mathbf{R}\}$). Note that, while the proposed discrete heuristics are not necessarily aligned, the continuous ones are nested (Table 3.1). Therefore, while discrete heuristics can be directly compared through their performance gaps, continuous strategies need to be assessed by the marginal performance gap obtained as compared to heuristics in the lower levels. Also, note that the proposed heuristics require different amounts of information for their implementation. Some require the details of the network structure and/or knowledge about the opponent's strategy. Others have free parameters and their implementation requires an exploration of the parameter space, with computations of expected vote shares for evaluating each value of the parameters. Table 3.1 summarises what information is required for each of the heuristics along with the nested structure.

Fig. 3.7 illustrates results of the gap in vote share ΔX for experiments in three budget scenarios (advantage, equality, and disadvantage) and against a discrete passive controller who targets $K = 16$ random nodes in the network. Note that we make a clear separation between the discrete (*purple*) and continuous (*orange*) regimes and that their vote shares are compared to the discrete or continuous optimal strategies, respectively.

From Fig. 3.7, we make the following observations about the three budget scenarios. First, when in budget advantage (panel **a**), we see that shadowing performs the best among discrete heuristics (with $\Delta X = 1.58 \cdot 10^{-2}$) closely followed by the degree–based heuristic ($\Delta X = 2.41 \cdot 10^{-2}$). Among continuous heuristics, the uniform ($\Delta X = 2.14 \cdot 10^{-2}$) and degree–based ($\Delta X = 2.09 \cdot 10^{-2}$) heuristics perform clearly worse than shadowing ($\Delta X = 2.55 \cdot 10^{-3}$) and shadowing plus shielding

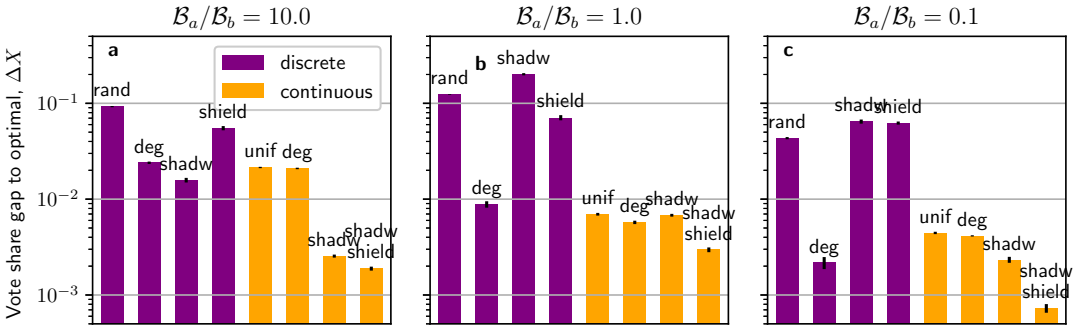


FIGURE 3.7: Comparison of various heuristics to optimal allocations on an email–interaction network and when the opponent targets $K = 16$ random nodes in the network in a discrete fashion. Bars represent the gap in vote share ΔX of the heuristic with respect to optimal numerical allocations for three different budget scenarios (as indicated on the top of the panels). Labels on top of bars refer to the following heuristics: random (*rand*), degree–based (*deg*), shadowing (*shadw*), shielding (*shield*), uniform (*unif*), and shadowing plus shielding (*shadw shield*). Error bars represent standard errors for 15 instances of the experiments.

($\Delta X = 1.89 \cdot 10^{-3}$). We can conclude that, when in budget advantage, shadowing is an important driver in both the discrete and continuous regimes. Second, for budget equality (panel **b**), the degree–based heuristic ($\Delta X = 8.99 \cdot 10^{-3}$) performs the closest to the optimal discrete strategy. Among the continuous heuristics, shadowing plus shielding is again most effective ($\Delta X = 2.97 \cdot 10^{-3}$), with the other three heuristics within the range $\Delta X \in [5.5 \cdot 10^{-3}, 7.0 \cdot 10^{-3}]$. We see that, when in budget equality, node degree is a very important factor in the discrete regime while shadowing plus shielding is most effective in the continuous regime. Last, when in budget disadvantage (panel **c**), the discrete degree–based heuristic ($\Delta X = 2.12 \cdot 10^{-3}$) again proves remarkably superior to the other heuristics, which do not perform better than random allocations. Among the continuous heuristics, shadowing plus shielding ($\Delta X = 7.28 \cdot 10^{-4}$) is again performing best, with the other three performing in the range $\Delta X \in [2.33 \cdot 10^{-3}, 4.44 \cdot 10^{-3}]$. In the budget disadvantage scenario, the degree–based heuristic is again a very important factor in the discrete regime, while shadowing plus shielding is predominant in the continuous regime. An alternative assessment of the heuristics based on computational efforts is shown in Appendix A.10.

In summary, in the discrete regime, node degree is consistently found to be the main driver for optimal allocations —only slightly surpassed by shadowing when in budget advantage. In contrast, in the continuous regime, shadowing and shielding have a considerably stronger effect than node degree across all scenarios. This effect is particularly pronounced when the active controller is in budget advantage —where shadowing is very effective or in budget disadvantage —where shielding is dominant. Reproductions of experiments from Fig. 3.7 in other network topologies can be found in Figs. A.14 and A.15 from Appendix A.8, showing qualitatively similar results.

3.9 Optimal strategies when both controllers are active

Up to this point, we have considered a passive opponent who holds a fixed strategy that is known by the active controller. In this section, we investigate the case where both controllers actively optimise their strategy. We assume that they simultaneously choose their strategy without having any information about the opponent's. However, since optimal strategies strongly depend on the opponent, we are interested here in finding a pair of strategies that provide no regret to either controller —i.e. the *pure–strategy Nash equilibrium* (Osborne, 1994)— so neither of them is willing to deviate from it.

We can explore first the Nash equilibrium from the analytical first–order solutions obtained in Sec. 3.5. Based on Eqs. (3.12) and (3.14), and exploiting the symmetry in exchanging the A and B controllers (as they are both active now), we can find mutually optimal responses as $w_{ai} = \mathcal{B}_a/N,\ w_{bi} = \mathcal{B}_b/N$ —i.e. uniform targeting. However, it is worth noting that this equilibrium point is not necessarily stable[3]. In fact, it can be proven that in the low–budget limit, uniform allocation is an unstable equilibrium point if the ratio between budgets exceeds $3 + 2\sqrt{2}$.

Alternatively, the Nash equilibrium can be assessed from analytical first–order approximations by looking at the expression of the total vote shares X in the limiting cases of low and high allocation strength with respect to node degree (X^L and X^H, respectively). Given a passive controller, we can consider how exploitable such controller is depending on its influence allocation profile $\boldsymbol{w}_b$. For the low–allocation limit, this would be

$$X^L = \frac{\mathcal{B}_a}{\mathcal{B}_a + \mathcal{B}_b} + \frac{1}{4\mathcal{B}_b} \sum_i \frac{(w_{bi} - \langle w_{bi} \rangle)^2}{d_i} \,, \tag{3.15}$$

where $\langle w_{bi} \rangle = \mathcal{B}_b/N$. We note that the vote share X^L obtainable for the active controller is the higher the more the allocations of the passive controller w_{bi} deviate from the mean $\langle w_{bi} \rangle$. In other words, the passive controller is the more exploitable by an optimal controller the more her passive strategy deviates from uniform targeting, and this relation is quadratic. Furthermore, the impact of this deviation is the stronger the lower the degree d_i of the node mis-targeted; thus, contrary to common intuition, controllers are the more exploitable the less carefully they allocate resources to low–degree nodes. We numerically confirm this result via a set of experiments on the email–interaction network and a passive opponent who targets all with strength $\langle w_{bi} \rangle = \mathcal{B}_b/N$ except for 40 of them, 20 receiving zero allocations and 20 receiving double allocations ($2\langle w_{bi} \rangle$). These 40 perturbations are only applied to nodes of a

[3]Stability in this context can be understood from how controllers would repeatedly react after a perturbation from the Nash equilibrium point. If the iterations of reactions after the perturbation asymptotically drives strategies back to the Nash equilibrium point, it can be said that it is stable.

specific degree d_p. Fig. 3.8 shows the deviations in vote share ΔX (with respect to the Nash equilibrium point) when such perturbations are applied to different degrees d_p in the network. Analytical and numerical results match well, with deviations in vote share ΔX decreasing with the node degree of the perturbations as $\Delta X \propto d_p^{-1}$ as predicted in Eq. (3.15).

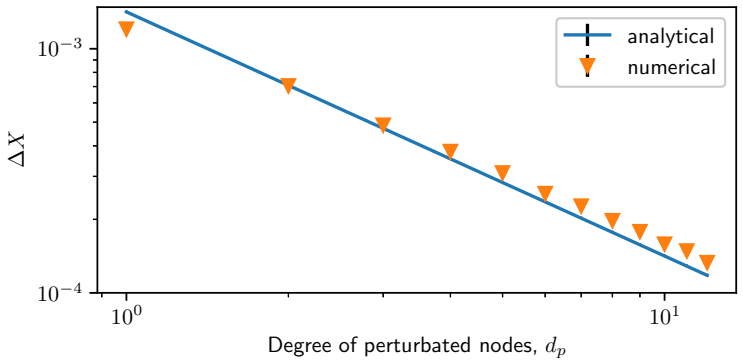


FIGURE 3.8: Increase in vote share ΔX for an optimal controller when the passive opponent deviates from uniform allocation on 40 nodes with degree d_n. Error bars represent standard errors over 15 instances of the experiment and are smaller than the symbols.

Conversely, the expression of the vote share X^H in the high–allocation limit is

$$X^H = 1 - \frac{\mathcal{B}_b}{\mathcal{B}_a + \mathcal{B}_b} \left(\frac{1}{N} \sum_i \sqrt{w_{bi} / \langle w_{bi} \rangle} \right)^2 . \tag{3.16}$$

Again, we find that the vote share X^H increases with deviations of the passive opponent from uniform allocations $w_{bi} = \langle w_{bi} \rangle$. In this limit, we find no effect of node degree d_i, i.e. controllers are equally exploitable independent of on which nodes they misallocate resources.

Lastly, uniform targeting at the Nash equilibrium can also be interpreted from the perspective of the shielding heuristic. A shielding strategy gains more advantage the more an opponent concentrates their influence on a few nodes, as shown in Fig. 3.5. Consequently, spreading allocations smoothly across all nodes in the network prevents controllers from being exploited by the opponent via shielding.

3.10 Discussion

In this chapter, we have studied influence maximisation on the voter model from the perspective of external controllers that influence nodes in the network via building links to them. In contrast to most previous literature, which assumes discrete settings of control, we have presented continuous control allocations. To address this problem,

we have developed a numerical optimisation approach based on gradient ascent, which can achieve solutions with arbitrary accuracy and in polynomial time.

Additionally, we have explored the structure of optimal allocations in influence maximisation for the voter dynamics. Our focus has been on scenarios with two opposing controllers, one passive and one actively optimising, for which we have defined the continuous influence maximisation problem. As an initial result, we have seen that the continuous regime adds some improvement in the objective function over the discrete regime, which can be linked to the benefits of spreading control allocations across the whole network. Regarding the structure of optimal allocations, we have pointed to three important heuristics that can be used for its characterisation: shadowing, shielding, and degree dependence. Despite the three heuristics playing a role in explaining optimal allocations in both the continuous and discrete regimes, their relative contributions vary. While optimal allocations in the discrete regime are mainly driven by node degree, the shielding and shadowing heuristics are predominant explanatory factors in the continuous regime. In the game–theoretical scenario where both controllers simultaneously optimise their influence allocations, the pure-strategy Nash equilibrium in the continuous regime is assumed when both controllers uniformly target all nodes in the network —regardless of details of the network topology.

Our results mainly apply to the scenario in which a controller reacts against a passive opponent. However, note that this framework can be mapped to the scenario in which a single controller attempts to influence a population that manifests some resistance to the external influence, as shown in Brede et al. (2018). In particular, such resistance to external control is equivalent to a passive opponent who targets nodes with a strength proportional to the degree of the node, so a degree–dependent strategy becomes much more predominant when only one external controller is featured.

This chapter contributes to the general understanding of controlling dynamical processes on networks (Liu and Barabási, 2016) in three ways. First, it provides three explanatory factors as tools that can be employed to characterise any solution to a control problem: degree dependence, shadowing, and shielding. Second, it shows that, in the particular case of the voter model with external controllers, the degree–dependence factor is the main driver of controlling solutions in the discrete regime, while the shadowing plus shielding combination is the main explanatory factor behind solutions in the continuous regime. Note that our mathematical framework for the voter model is fully mappable to the DeGroot dynamics with zealots or to some instances of the Friedkin–Johnsen, so all results here shown are directly translated to those models. For other opinion models, these results can be used for guidelines as to what to expect. Lastly, the proposed factors can be employed as heuristics in cases where optimal solutions cannot be easily obtained numerically, or alternatively as benchmarks for testing approximately optimal solutions.

Some of our results relate to other findings from previous literature. Most works refer to a preference towards hubs when linking optimal solutions to a degree dependence. However, we have shown that correlations with node degree can also be negative (i.e. targeting peripheral nodes), particularly when the controller is in budget disadvantage). Brede et al. (2019a, 2018, 2019b) also show that the preference towards hub nodes shifts to a preference towards periphery nodes in the discrete regime, although under other conditions; namely when optimising for short time horizons (Brede et al., 2019a) or in the presence of noise in the opinion dynamics (Brede et al., 2018, 2019b). They also suggest that, for short time horizons, the optimal strategy is also opponent–dependent, shadowing nodes with initially opposing opinions if they are few and avoiding them when they are many (Brede et al., 2019a). Our results have pointed to a lack of degree dependence in the continuous case. This independence is not fully reflected in the studies on the continuous IM on the Ising model with noise by Lynn and Lee (2017, 2018). Unlike in our results, they find that optimal targets are strongly related to node degree in the continuous regime, as periphery nodes should be targeted when levels of noise are low and hub nodes for intermediate levels. However, when noise levels are high, a uniform allocation of influence is the best strategy against an also uniform opponent, and hence there is no correlation with node degree (Lynn and Lee, 2017). In light of these results, studying the inclusion of noise would be a natural extension to this work, as it could also generate a degree dependence in the continuous IM for the voter dynamics. A related rationale is discussed by Yildiz et al. (2013) in their solutions to Influence Maximisation with discrete allocations. They show that optimal solutions obtained have a 'bottleneck property' in that they are placed to shield the opponent's influence to the rest of the network. However, their optimal allocations are also placed as far as possible from the opponent's, while shielding favours the direct neighbours of the opponent's targets.

There are many other ways in which this work can be extended. For instance, we use in our experiments a heterogeneous network that is generally well mixed —without strong communities— and has very little degree assortativity —i.e. correlations in node degree across edges. It would be of interest to study the impact of these topological modifications on shadowing and shielding. A further limitation of our work is that we have only considered strategies that do not vary over time and optimisation is set for the long–term steady state. However, campaigners in more realistic settings might care to achieve their goals in a limited time, and could probably also vary their strategy at different stages of the dynamics. For an extension to influence maximisation with dynamic allocations, see e.g. Cai et al. (2021); Brede et al. (2019a). Furthermore, the structure of the influence network may be uncertain or dynamic or the opponent may also adapt their strategy. This added complexity requires a different approach towards influence maximisation, able to cope with uncertainties by adapting the controller's strategy. Reinforcement learning techniques

may be appropriate for these contexts (Tran-Thanh et al., 2021; Mandel and Venel, 2020; Ali et al., 2020).

Additionally, the gradient ascent algorithm can be further improved for a better performance and scalability. The exact computation of the expected steady–state on the voter model requires the inversion of a matrix whose size is equal to the number of nodes in the system, which scales as $O(N^3)$, and must be done at every step of the gradient ascent algorithm. A possible improvement involves performing this computation numerically via Jacobi iterations, as the matrices to be inverted are diagonally dominant (Masuda, 2015), which would speed up the algorithm in very large networks. A different means to reduce the time complexity of the algorithm is computing gradients based on the heterogeneous mean–field approximation instead of the exact formulation. This modification would eliminate the matrix inversion altogether and drastically reduce computation time at the expense of accuracy, particularly for networks where this approximation is shown to perform worse (i.e. sparse, heterogeneous networks). For this case, it would also be advisable to further test the accuracy of the approximation for other conditions, such as networks with high degree assortativity or strong community structure. Alternatively, results from the mean–field approximation could be used as seeds for the Jacobi iterations, which would significantly reduce the number of iterations while being robust to small deviations of the mean–field approximation. The step size of the gradient ascent algorithm could also be chosen in a more principled manner. More theoretical backing can be given for choosing a step size that guarantees convergence of the algorithm, or automatic methods other than backtracking for dynamically adapting the step sizes could be developed. Additionally, the algorithm could be improved by employing the Hessian of the optimisation function for performing steps instead of the gradient (Goodfellow et al., 2016).

In conclusion, we have highlighted the impact that having a continuous or discrete regime may have on both the quality and structure of optimal solutions to network control. We have provided a methodology to arrive at solutions to the problem plus three factors that can be used to characterise solutions to control diffusion problems where opposing external controls influence the network: shadowing, shielding, and degree dependence. The application of this framework to other contexts may facilitate comparisons across different modelling decisions and problems.

Chapter 4

Bias and Influence Maximization: When to Target Biased Agents?

This chapter is mostly based on a paper published in the conference proceedings of the conference CompleNet2020 (Romero Moreno et al., 2020a). All the writing from this chapter (and the original paper) is mine.

4.1 Introduction

The previous chapter investigates the structure of optimal influence on homogeneous populations that are connected through complex networks. However, as pointed out by Aral and Dhillon (2018), past approaches to Influence Maximisation often overlook aspects of heterogeneity in agent behaviour. Specifically, in the context of the Independent Cascade model, it has been shown that results of IM are strongly affected by agents' susceptibility to adopt opinions (Aral and Dhillon, 2018).

In the context of the Voter Model, previous work has already considered a heterogeneous presence of biases in the adoption of opinions. Biased agents can be referred to as *zealots*, although this may cause some confusion with the more–common interpretation that 'zealots' refer to inflexible agents that never change their opinion (Mobilia, 2003; Mobilia et al., 2007). A more nuanced approach could consider agents that tend to adopt some of the opinions more easily than others, which typically accounts for personal biases or external influences not captured in a typical model. For a clear distinction, this type of agents can be called *biased agents* or *partial zealots* (Masuda et al., 2010).

While Aral and Dhillon (2018) demonstrate that the amount of bias influences optimal allocations in the IM, they have not explored detailed mechanisms for such differences. Here, we extend the analysis of optimal allocations of influence to a more

complex and realistic scenario where agents may hold different levels of bias towards one of the opinions. An optimal campaign manager with limited resources would never target perfect zealots whose opinions cannot be influenced. However, under what conditions would she target biased agents and how does this relate to these agents' topological position in the social network?

Below, we shall explore these questions for different network topologies. As biased agents represent partly radicalised agents, answers to the above questions might shed light on how to reduce radicalisation in social systems (Vendeville et al., 2022). Our analysis, based on both analytical and numerical results, shows a rich diagram of preferences and dependencies of allocations to biased and unbiased agents, on degree, which depend on and vary with the controller's budget. Our results emphasize that heterogeneity in agent properties strongly affects strategies for IM on heterogeneous networks.

The chapter is organised as follows. Section 4.2 includes the presence of bias in the formulation of the continuous IM problem. Section 4.3 introduces the experimental settings and the methodologies employed. Sections 4.4, 4.5, and 4.6 present analytical and numerical solutions to the problem with a single external controller on simple graph topologies, and later extends the analysis to heterogeneous Barabasi–Albert networks. Section 4.7 expands the analysis to the case of two external controllers and bias against either opinion. Last, section 4.8 discusses and summarises our main conclusions.

4.2 Model

Departing from the basic problem formulation of IM on the VM (Sect. 3.2), we introduce the following modifications. Bias against opinion A is modelled by an intrinsic level of bias b_{Ai} that gives the probability for agent i of not copying opinion A when this should happen according to the rules of the Voter Model (Masuda et al., 2010). Analogously, an intrinsic level of bias b_{Bi} gives the probability that agent i does not copy opinion B. We assume here that agents can only be biased against one of the opinions; agents biased against both opinions simultaneously would reflect a level of stubbornness (Nishi and Masuda, 2013; Martins and Galam, 2013; Pérez-Llanos et al., 2020), a mechanism slightly different to that of bias, which will be the focus of the chapter.

For this scenario with biased agents, the rate equations (3.1) for the probability x_i of node i to hold opinion A are re-defined as

$$\frac{dx_i}{dt} = (1 - b_{Ai})\,(1 - x_i)\frac{\sum_j w_{ij}x_j + w_{ai}}{d_i + w_{ai}} - (1 - b_{Bi})x_i\frac{\sum_j w_{ij}(1 - x_j) + w_{bi}}{d_i + w_{ai}}\,, \tag{4.1}$$

where $W = (w_{ij})$ is the weighted adjacency matrix and $d_i = \sum_j w_{ij}$ is the degree of node i. Note that the left term on the right-hand side of the equation accounts for the probability of a node holding opinion B and copying the opinion of a neighbour with A (which is reduced by the level of bias b_{Ai}), while the right term reflects the probability of a node holding opinion A and copying the opinion of a neighbour with B (which is reduced by the level of bias b_{Bi}).

We again focus on IM in the expected equilibrium state, which is unique and asymptotically reached irrespective of initial conditions (Yildiz et al., 2013). The probabilities x_i^* of adopting opinion A in the steady–state can be determined from $dx_i/dt = 0$, leading to the system of N second–order equations

$$0 = (1 - b_{Ai}) \sum_j w_{ij} x_j^* + (b_{Ai} - b_{Bi})\, x_i^* \sum_j w_{ij} x_j^* - \\ - x_i^* \left[(1 - b_{Bi})(d_i + w_b) + (1 - b_{Ai})\, w_{ai}\right] + (1 - b_{Ai})\, w_{ai}\,. \quad (4.2)$$

Since this system of equations is hard to solve numerically, vote shares in the steady–state can alternatively be found by numerical integration of Eq. (4.1).

As we lack an explicit expression for the total vote share $X = \sum_i x_i^*/N$ in the steady–state for general networks, we resort to numerical methods for the optimisation process. Similarly to the basic formulation of the continuous IM, gradients can also be computed after the inclusion of biases, so the gradient–ascent technique from Sec. 3.3.1 is also applicable here. The gradient $\nabla_{w_a} X$ can be computed by first applying partial derivatives to Eq. (4.2),

$$0 = (1 - b_{Ai}) \sum_j w_{ij} \frac{\partial x_j^*}{\partial w_{ak}} + (b_{Ai} - b_{Bi}) \left(\frac{\partial x_i^*}{\partial w_{ak}} \sum_j w_{ij} x_j^* + x_i^* \sum_j w_{ij} \frac{\partial x_j^*}{\partial w_{ak}} \right) - \\ - \frac{\partial x_i^*}{\partial w_{ak}} \left[(1 - b_{Bi})(d_i - w_{bi}) + (1 - b_{Ai})\, w_{ai}\right] + \delta_{i,k}(1 - b_{Ai})(1 - x_i^*)\,, \quad (4.3)$$

where $\delta_{i,k}$ is the Kronecker delta, with value 1 when $i = k$ and zero otherwise. Eq. (4.3) is a system of linear equations whose solution gives us the final gradient as

$$\nabla_{\boldsymbol{w}_a} X = \frac{1}{N} (\nabla_{\boldsymbol{w}_a} \boldsymbol{x}^*)^T \mathbf{1} = \frac{1}{N} (I - B_A)\, \mathrm{diag}(1 - x_i^*) \left[(B_A - B_B)\, \mathrm{diag}\left(x_i^* - \sum_j w_{ij} x_j^* \right) + \right. \\ \left. + (I - B_B)(D + W_b) + (I - B_A)(W_a - W) \right]^{-1} \mathbf{1}\,, \quad (4.4)$$

where L is the Laplacian of the network, B_A, B_B, and W_a are diagonal matrices with values $B_{A(ii)} = b_{Ai}$, $B_{B(ii)} = b_{Bi}$, and $W_{ii} = w_{ai}$, respectively, bold symbols are vertical vectors, and the values of x_i^* are computed via numerical integration, as explained above.

4.3 Methods and experimental settings

For the sake of simplicity, we will first consider in Secs. 4.4, 4.5 and 4.6 that there only is a single external controller (the one favouring opinion A, without loss of generality) and that agents in the population hold a bias against the controller's opinion; i.e. we are interested in optimal strategies for the controller to overcome the population bias. In Sec. 4.7, we will look at the general case having two external influences — an active controller and a passive opponent — and agents that can be biased against either opinion. We consider the population of biased agents against an opinion to be a minority and hence we set their numbers to 20% of the total population.

To gain progressive understanding about the role of bias in influence maximisation in the VM, we examine experiments on networks with increasing complexity. First, in Sec. 4.4, we explore optimal allocations on the complete graph, i.e. without having a topology at all. This basic experiment will allow us to analytically obtain closed–form solutions to understand the effect of budget availability and the level of bias on optimal strategies in the thermodynamic limit of infinitely large networks ($N \to \infty$). These analytical results are contrasted with numerical experiments in finite networks obtained via the gradient ascent algorithm presented in Chapter 3.

Next, in Sec. 4.5, we perform experiments on a bipartite graph. This graph topology allows to directly study the role of node degree on optimal allocations while removing the effect of higher–order relations — i.e. neighbours of neighbours. While closed–form solutions of optimal allocations cannot be obtained for this graph topology, numerical solutions for infinitely large networks can still be obtained easily via a numerical solver for polynomial equations. These results are also contrasted with numerical experiments on finite networks obtained via the gradient ascent algorithm.

In Sec. 4.6, we further increase the complexity of the network topology by performing experiments on networks with much more heterogeneous degree distributions — i.e. synthetic Barabasi–Albert networks — whose network structure resembles more the ones found in real social systems. As the complexity of this scenario is much bigger, we can only obtain numerical results via the gradient ascent algorithm for this network topology.

Last, in Sec. 4.7, we study a more general scenario that includes a passive opponent and three types of agents: unbiased agents, agents biased against A, and agents biased against B. As the space of parameters greatly increases, we only study the simplest network topology — the complete graph — and analyse numerical results via gradient ascent for this scenario.

In all these experiments, numerical integration to find the steady–state was performed down to a precision in X of 10^{-5}, with an exponentially increasing time step, $\Delta t = 1.5^t$.

The gradient ascent was performed for a maximum of 10^3 iterations and an initial learning rate $\mu = 500$. Complete and bipartite graphs are of size $N = 100$, as their homogeneity limits the effect of the system size. Barabasi–Albert networks are of size $N = 1000$ for general explorations of the parameters and of size $N = 5000$ for analysing the dependency of allocations on node degree. Barabasi–Albert networks were generated following preferential attachment rules, with every new node linking to two existing ones — resulting in an average degree of $\langle d \rangle \approx 4$. All the code used for performing these experiments and producing the figures is available at `https://git.soton.ac.uk/grm1g17/zealots`.

4.4 Complete graph: the roles of the allocation budget and the level of bias on optimal allocations

We start our analysis with an infinitely large complete graph where a fraction ρ_A of nodes are biased agents with equal bias b_A against opinion A, and the rest of the nodes are normal (unbiased) agents with $b_{Ai} = b_{Bi} = 0$. We assume that there is a single external controller that allocates the same link weight to agents of the same type, where w_{aA} and w_{an} are the weights of link allocations given to biased or normal agents, respectively, and $\hat{w}_{aA}$ and $\hat{w}_{an}$ express these link weights as fractions with respect to the incoming links from other peers in the network; i.e. $\hat{w}_{aA} = w_{aA}/N$ and $\hat{w}_{an} = w_{an}/N$. Analogously, since the network has infinite size instead of working with an (infinite) absolute budget $\mathcal{B}_a$, we refer to budget densities, $\hat{\mathcal{B}}_a = \mathcal{B}_a/N^2$, that represent allotted per-node allocations with respect to the system size *if the budget was evenly spread across all nodes*. Note that, since in the complete graph every node is connected to every other node in the network, $\hat{\mathcal{B}}_a = 1$ implies a budget power that allows the controller to target all nodes in the network so that agents select the external controller 50% of the time when following the voting dynamics.

For this scenario, general rate equations from Eq. (4.1) transform into

$$\begin{aligned}
\frac{dx_A}{dt} &= (1-b_A)\,(1-x_A)\frac{\rho_A\,x_A + (1-\rho_A)\,x_n + \hat{w}_{aA}}{1+\hat{w}_{aA}} - x_A\frac{1-\rho_A\,x_A-(1-\rho_A)\,x_n}{1+\hat{w}_{aA}}\,,\\
\frac{dx_n}{dt} &= (1-x_n)\,\frac{\rho_A\,x_A + (1-\rho_A)\,x_n + \hat{w}_{an}}{1+\hat{w}_{an}} - x_n\frac{1-\rho_A\,x_A-(1-\rho_A)\,x_n}{1+\hat{w}_{an}}\,.
\end{aligned} \tag{4.5}$$

The unique decision parameter for the distribution of budget among both types of agents can be defined by the fraction $\alpha_A \in [0,1]$ of budget allocated to biased agents, such that

$$\begin{aligned}
\hat{\mathcal{B}}_a &= \alpha_A\hat{\mathcal{B}}_a + (1-\alpha_A)\hat{\mathcal{B}}_a = \rho_A\,\hat{w}_{aA} + (1-\rho_A)\hat{w}_{an},\\
\hat{w}_{aA} &= \alpha_A\hat{\mathcal{B}}_a/\rho_A,\\
\hat{w}_{an} &= (1-\alpha_A)\hat{\mathcal{B}}_a/(1-\rho_A)\,.
\end{aligned}$$

The whole influence budget is then focused on normal agents when $\alpha_A = 0$, on biased agents when $\alpha_A = 1$, and link weights are equal for both types of agents when $\alpha_A = \rho_A$.

The total vote share in the equilibrium for this scenario can be easily derived from Eq. (4.5) as

$$X = \rho_A\, x_A^* + (1-\rho_A)\, x_n^* = \frac{\hat{\mathcal{B}}_a}{b_A\,\rho_A}\,\frac{\alpha_A(1-b_A)\hat{\mathcal{B}}_a(1-\alpha_A) + (1-\rho_A)\rho_A(1-\alpha_A\, b_A)}{\hat{\mathcal{B}}_a(1-\alpha_A) + (1-\rho_A)\rho_A}\,, \tag{4.6}$$

with a boundary at $X = 1$. From Eq. (4.6), we can see that the vote share generally increases with budget availability $\hat{\mathcal{B}}_a$ and decreases with the fraction of biased agents ρ_A and their bias b_A. Optimal allocations α_A^* can be found by solving $\partial X/\partial \alpha_A = 0$, arriving at

$$\alpha_A^* = 1 - \frac{\rho_A(1-\rho_A)}{\hat{\mathcal{B}}_a}\left(\frac{1}{\sqrt{1-b_A}} - 1\right)\,, \tag{4.7}$$

which is bounded at $\alpha_A^* = 0$. This result shows that, for a given ρ_A, optimal allocations target biased agents the more the bigger the budget $\hat{\mathcal{B}}_a$ and the smaller their bias b_A. We can also find the switching point of bias b_A^* at which normal agents start to be favoured over biased agents (i.e. when $\alpha_A^* = \rho_A$),

$$b_A^* = 1 - \left(\frac{1}{\hat{\mathcal{B}}_a/\rho_A + 1}\right)^2\,, \tag{4.8}$$

which increases with the budget $\hat{\mathcal{B}}_a$ and decreases with the density of biased agents ρ_A. So a controller with high budget can afford to target biased agents with a higher bias, and especially when there are fewer of these in the population.

We compare the analytical solutions developed above with numerical results obtained via gradient ascent as introduced in Sec. 4.2. For this purpose, we analyse how optimal allocations to biased agents α_A^* vary with the level of bias b_A and budget density $\hat{\mathcal{B}}_a$ (Fig. 4.1**a**) and observe the resulting equilibrium vote shares X (Fig. 4.1**b**). We note that for low bias b_A or high budget density $\hat{\mathcal{B}}_a$, the controller achieves full control $X = 1$ by both analytical and numerical methods (c.f Fig. 4.1**b**). We do not show optimal allocations from the gradient ascent algorithm when $X = 1$ for the sake of clarity, as in those cases a range of possible allocations already reach the maximum vote share $X = 1$ and numerical results —sensitive to the initialisation of the algorithm— can be noisy. Conversely, analytical optimisations do not take into account the boundary at $X = 1$ and provide the most robust strategy. For higher values of b_A or lower budget density $\hat{\mathcal{B}}_a$ (where full control is not possible and $X < 1$), allocation strategies from both methods are found to be in perfect agreement, as well as the equilibrium vote shares X achieved by them.

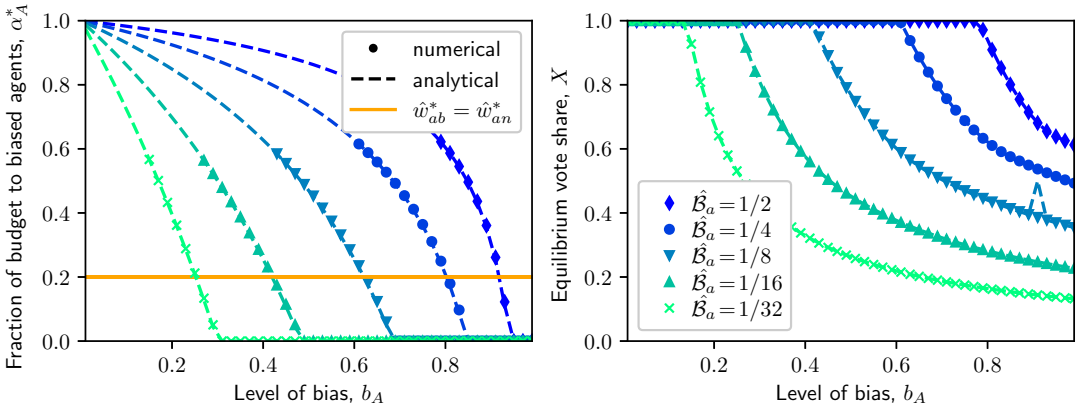


FIGURE 4.1: **a** Optimal fraction of budget given to biased agents α_A^* and **b** their resulting vote shares in the equilibrium X on a complete network with a fraction of biased agents equal to $\rho_A = 0.2$ and for varying levels of bias b_A and varying budget density $\hat{\mathcal{B}}_a = \mathcal{B}_a/N^2$. *Dashed lines* give results obtained via analytical methods (on an infinite graph) while *symbols* give results obtained via numerical methods (on graphs of size $N = 100$). Optimal influence allocations from numerical methods are omitted when $X = 1$ for improved clarity of the graph. The *orange, horizontal line* is placed at $\hat{w}_{aA}/\hat{\mathcal{B}}_a = \alpha_A^*/\rho_A = 1$, and represents equal influence weights given to both types of agents.

In general, optimal strategies favour allocations to biased agents over normal agents for low values of b_A and high values of $\hat{\mathcal{B}}_a$. This behaviour gradually decreases with b_A, switching to favouring normal agents at some critical bias level b_A^* (in Fig. 4.1**a**, crossing the horizontal orange line) and fully avoiding biased agents ($\alpha_A^* = 0$) at another critical bias level b_A^{**}. Both critical points depend on the budget density $\hat{\mathcal{B}}_a$. Equilibrium vote shares reach full control ($X = 1$) for low values of b_A, experience a steep drop just after leaving the full control regime, and gradually decelerate in their decrease as b_A increases.

4.5 Bipartite graph: the role of node degree on optimal allocations

Next, as a way to explore the interplay of node degree and bias, we consider complete bipartite graphs, i.e. graphs with nodes divided into two groups where nodes from a group are connected to all nodes in the other group with no intra-group links. We assume that graphs are infinitely large and that nodes in the smaller group (which we will refer to as *hubs*) comprise a fraction $\rho_h = 0.2$ of the total network and have a bias $b_{Ah} = b$, while nodes from the larger group (*periphery*) have a complementary bias $b_{Ap} = 1 - b$. Analogous to the groups defined above for the complete graph, here we assume that the external controller allocates the same link weight to agents of the same type, with w_{ah} and w_{ap} referring to the link weights connecting each hub or

periphery node, respectively, to the controller, and $\hat{w}_{ah}$ and $\hat{w}_{ap}$ give these weights as fractions with respect to the number of nodes in the network —i.e. $\hat{w}_{ah} = w_{ah}/N$ and $\hat{w}_{ap} = w_{ap}/N$. Rate equations for this case correspond to

$$\begin{aligned} \frac{dx_h}{dt} &= (1-b)\,(1-x_h)\frac{(1-\rho_h)\,x_p + \hat{w}_{ah}}{1+\hat{w}_{ah}} - x_h\frac{(1-\rho_h)\,x_p}{1+\hat{w}_{ah}}\,, \\ \frac{dx_p}{dt} &= b\,(1-x_p)\,\frac{\rho_h\,x_h + \hat{w}_{ap}}{1+\hat{w}_{ap}} - x_p\frac{\rho_h\,x_h}{1+\hat{w}_{ap}}\,. \end{aligned} \tag{4.9}$$

We use now the controller's decision parameter as the fraction $\alpha_h \in [0,1]$ of budget given to hub nodes,

$$\hat{w}_{ah} = \alpha_h\hat{\mathcal{B}}_a/\rho_h, \quad \hat{w}_{ap} = (1-\alpha_h)\hat{\mathcal{B}}_a/(1-\rho_h)\,,$$

leading to a vote share in the equilibrium equal to

$$X = \frac{\hat{\mathcal{B}}_a}{b\,\rho_h}\frac{(1-b)[\alpha_h\hat{\mathcal{B}}_a(1-\alpha_h)+\rho_h(1-\rho_h)]}{(1-\rho_h)^2} + \frac{(1-b)\alpha_h\hat{\mathcal{B}}_a(1-\alpha_h)+\rho_h(1-\rho_h)(\alpha_h b+1)}{\hat{\mathcal{B}}_a(1-\alpha_h)+\rho_h(1-\rho_h)}\,, \tag{4.10}$$

with again a boundary at $X = 1$. Optimal allocations α_h^* can again be found by solving $\partial X/\partial\alpha_h = 0$, which results in a fourth-order polynomial equation that we evaluate numerically.

Figure 4.2 analyses optimal allocations given to hubs on a bipartite network for the levels of bias defined above, obtained both by analytical and numerical methods. Figure 4.2**a** gives the fraction α_h^* of budget allocated to hub nodes for different levels of bias b and budget density $\hat{\mathcal{B}}_a$. We note that, for a large budget density (corresponding to *diamonds* in the figure), there is a consistent preference in allocations towards hub nodes (i.e. $\alpha_h^* > \rho_h = 0.2$) regardless of which group holds a higher bias —except in the limit of hubs being almost perfect zealots ($b_{Ah} \approx 1$). On the contrary, when the budget density is low (*crosses*), optimal allocations quickly switch from fully targeting hubs ($\alpha_h^* = 1$) to fully targeting periphery nodes ($\alpha_h^* = 0$), even when the bias of periphery nodes is larger than that of hubs —e.g. when $b_{Ah} = 0.4$ and $b_{Ap} = 0.6$. Figure 4.2**b** gives the resulting equilibrium vote shares X for optimal allocations of Fig. 4.2**a**. We note that, for low budgets (*crosses*), the vote shares obtained in optimal allocations are the higher the more the bias is concentrated on hubs. In contrast, for high budgets (*diamonds*), moderate bias in both groups ($b_{Ah} \approx b_{Ap} \approx 0.5$) leads to the highest vote shares from optimal allocations.

To summarize our findings on simple network topologies, we see that influence allocations to biased agents are preferred when their bias is low or the available budget is high, while allocations to normal agents are preferred otherwise. When nodes of different degrees are present, allocations to hubs are preferred over periphery

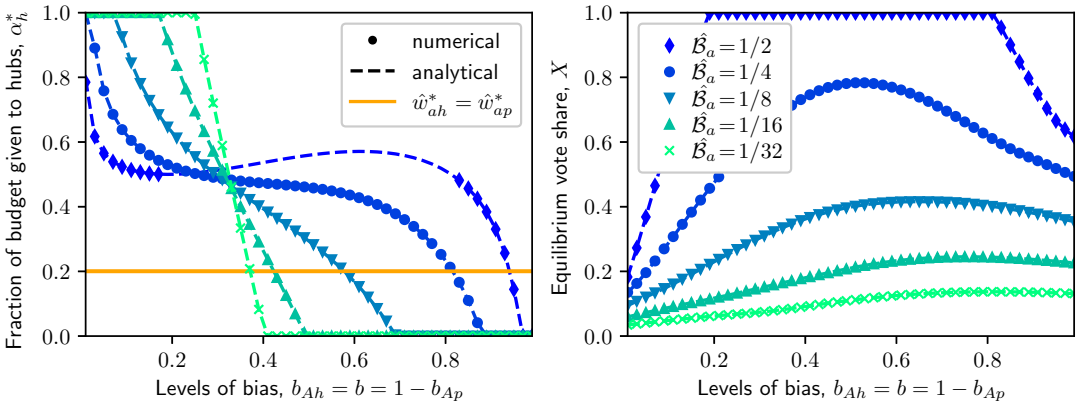


FIGURE 4.2: **a** Optimal fraction α_h^* of budget given to hub nodes and **b** resulting equilibrium vote shares X on bipartite graphs and for different levels of bias b and budget density $\hat{\mathcal{B}}_a$. *Dashed lines* give results obtained via analytical methods on infinite networks, while *symbols* give results obtained via numerical methods on graphs with size $N = 100$. Optimal influence allocations from numerical methods are omitted when $X = 1$ for improved clarity of the graph. Hub nodes, with bias b_{Ah}, compose $\rho_h = 0.2$ of the network, while periphery nodes compose the remaining 0.8 and have bias $b_{Ap} = 1 - b_{Ah}$.

nodes when the budget availability is high, even when the bias of hubs is higher than that of periphery nodes. In contrast, for low budgets, higher influence allocations to periphery nodes can be preferable even in cases when their bias is higher than that of hubs. So there is an interplay between bias —with a preference for biased agents proportional to budget availability— and degree —with a preference for hub nodes proportional to budget availability.

4.6 Barabasi–Albert networks: the role of degree heterogeneity

Next, we analyse results on Barabasi–Albert (BA) networks (Barabási and Albert, 1999) to further explore how the interplay of bias and node degree affects optimal targeting on heterogeneous networks with structures more akin to real social systems. The average degree in these networks is $\langle d \rangle \approx 4$, resulting in a much more sparse social system than with complete and bipartite graphs. Furthermore, most nodes have $d_i < 4$ while a few nodes may be connected to a large proportion of the population. As above, we induce a uniform bias b_A to a random fraction ρ_A of the population while keeping the remaining nodes as normal (unbiased) agents.

We first explore general behaviours by looking at average optimal allocations given to biased agents and the resulting equilibrium vote shares X (c.f Fig. 4.3). We note remarkable similarities between the results obtained here and those on the complete graph (Fig. 4.1) regarding both optimal allocations and equilibrium vote shares. Note

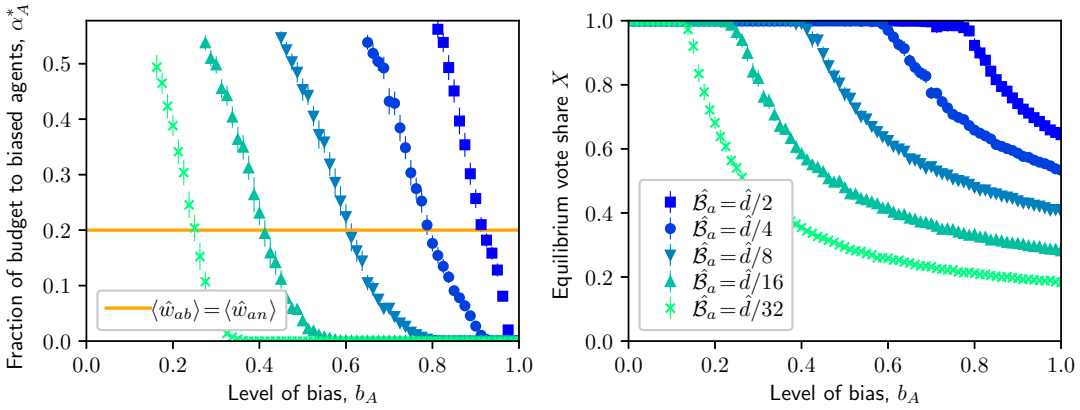


FIGURE 4.3: **a** Optimal fraction α_A^* of budget given to biased agents **b** and resulting equilibrium vote shares X on BA networks for varying bias b_A and varying budget density $\hat{\mathcal{B}}_a = \mathcal{B}_a / N^2$. Networks are of size $N = 1000$, mean degree $\langle d \rangle \approx 4$, and with a random fraction $\rho_A = 0.2$ of biased nodes with bias b_A. Every point is the average over 50 realizations of the experiment, with error bars giving three standard deviations from the mean. Symbols are omitted for optimal allocations that lead to full control ($X = 1$). The horizontal line (*orange*) represents equal link weights given to both types of agents on average.

that these similar results are retrieved after a re-scaling of the budget density $\hat{\mathcal{B}}_a$ by $\hat{d} = \langle d \rangle / N$ —as $\langle d \rangle = N$ in the complete graph. This is related to the notion of *effective* budget as discussed by Chinellato et al. (2015). Again, targeting biased agents is preferred when the bias is low or the available budget is high. This preference slowly shifts towards normal agents as b_A increases, eventually assigning them higher influence links —at b_A^*— and even focusing the whole budget on them —at b_A^{**}.

We next analyse the relationship between optimal allocations and node degree. More specifically, we want to find whether the link preferences to biased or normal agents are uniformly held across nodes with different degrees. Figure 4.4 displays optimal allocations grouped by node degree $\langle w_{ai}^* \rangle_d$ for a given budget density $\hat{\mathcal{B}}_a = \langle d \rangle / 16$ and three different bias parameters $b_A = 0.3, 0.5, 0.9$. We note clear correlations between optimal allocations and node degree in most cases. When the bias is relatively low ($b_A = 0.3$, Fig. 4.4**a**), the external controller exhibits a clear allocation preference towards high–degree nodes, among both biased and normal agents. For an intermediate bias ($b_A = 0.5$, Fig. 4.4**b**) optimal controls omit allocations to biased agents that do not have a very low degree, while mildly targeting those with the lowest degrees. This behaviour relates to our findings for complete bipartite graphs above, where biased agents at the periphery were preferred over biased agents at hubs for low budgets and mild bias values. High–degree nodes are still preferred among normal agents in this scenario. Last, when biased agents are highly stubborn ($b_A = 0.9$, Fig. 4.4**c**), they remain untargeted regardless of their degree, while the correlation with node degree generally decreases for normal agents, with high-degree nodes also untargeted.

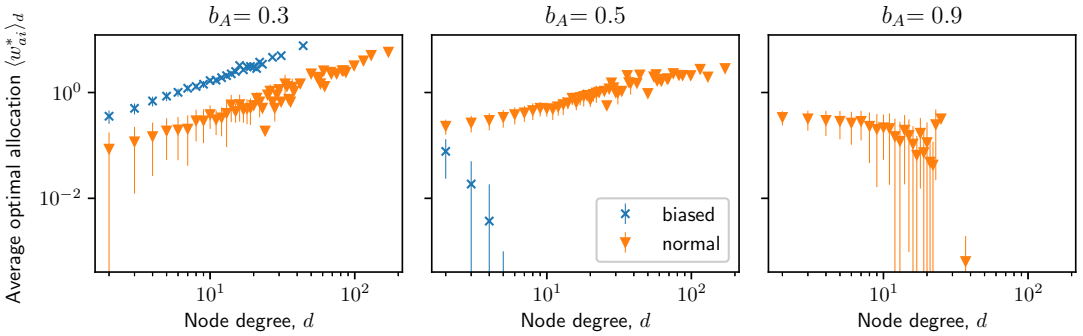


FIGURE 4.4: Average per-degree optimal allocations $\langle w^*_{ai}\rangle_d$ given to biased and normal agents depending on the node degree d and for three values of b_A on a BA-network of size $N = 5000$. Each point is the mean allocation over all nodes of same degree and error bars denote standard deviations. These results correspond to a single realization of the experiment. Note that allocations to normal agents in **c** are missing due to all values been zero.

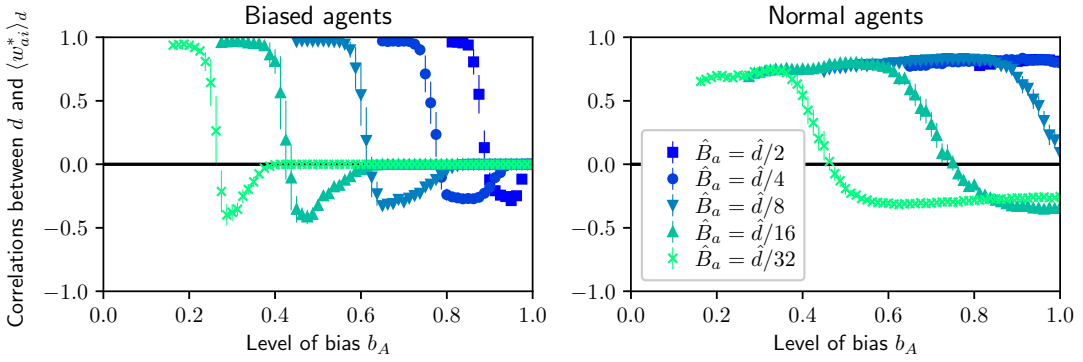


FIGURE 4.5: **a** Pearson correlations between per-degree optimal allocations $\langle w^*_{ai}\rangle_d$ and node degree d for biased and **b** normal (right) agents for different values of $\hat{B}_a$ and b_A on BA-networks of size $N = 1000$. Each point is the average correlation over 50 runs of the experiment with error bars accounting for three standard deviations from the mean, both computed in the Fisher transformation domain. Correlations are only shown for situations in which full control ($X = 1$) is not achieved. All p-values fall below 10^{-20}.

We extend the correlation analysis in Fig. 4.4 to other levels of bias and available budgets by measuring Pearson correlations between $\langle w^*_{ai}\rangle_d$ and d on BA-networks (Fig. 4.5). We note again clear patterns that are in agreement with Fig. 4.4: hubs are preferred when the bias is low, with the preference decreasing with b_A and switching to periphery nodes at some b^*_A. The switching points are lower for biased agents than for normal agents and increase with budget availability. Correlations on allocations to biased agents eventually reach zero, marking the point where they are fully untargeted.

Combining the information from average allocations given to biased agents (Fig. 4.3) and average per-degree allocations given to both groups (Fig. 4.5), we obtain the following general pattern. When the bias is low (but not enough for achieving full

control of the network), targeting biased agents is preferred over normal agents and hubs receive more allocation within each group. Preferences for biased agents —and especially those at hubs— diminish as the bias increases, going through three different transition points. The first transition point b_A^* marks a shift of preference towards normal instead of biased agents and a shift of preference within biased agents for those having a lower degree instead of hubs. After the second transition point b_A^{**}, no allocation is given to biased agents and the preference for hubs among normal agents starts to diminish. After the last transition point b_A^{***}, low–degree nodes are preferred over hubs among normal agents.

In summary, when networks with heterogeneous degree distributions are studied, a richer hierarchy emerges due to the interplay between node degree and agent type, with the following groups (sorted by their difficulty to be controlled in decreasing order): biased agents at hubs, biased agents at the periphery, unbiased agents at hubs, and unbiased agents at the periphery.

4.7 General model: Including a passive opponent and agents biased towards either opinion

Last, we analyse a more general scenario that also includes a passive opponent and in which there are three types of agents: agents biased against opinion A ($b_{Ai} > 0$), agents biased against opinion B ($b_{Bi} > 0$), and normal (unbiased) agents ($b_{Ai} = b_{Bi} = 0$). For this scenario, we only perform experiments on the complete graph, as to build intuition of the impact of an opponent in optimal allocations, and of agents biased against the opponent — i.e. prone to support the active controller's opinion — isolating these factors from the effect of the network topology. We assume that agents biased against opinion A and agents biased against opinion B constitute each a fraction $\rho_A = \rho_B = 0.2$ of the total population, with the remaining $\rho_n = 0.6$ being normal agents.

As above, we assume that controllers allocate the same link weight to agents of the same type. We again refer to budget densities to express the budget availability of the active controller ($\hat{\mathcal{B}}_a = \mathcal{B}_a / N^2$) and of its passive opponent ($\hat{\mathcal{B}}_b = \mathcal{B}_b / N^2$). Each controller now has two degrees of freedom when distributing their budget among the three types of agents, with $\alpha_A, \alpha_B, \alpha_n \in [0,1],\ \alpha_A + \alpha_B + \alpha_n = 1$, defining the fraction of the budget allocated by the active controller to each group, and $\beta_A, \beta_B, \beta_n \in [0,1],\ \beta_A + \beta_B + \beta_n = 1$, are the analogous choices of the passive opponent.

We start experiments by fixing the budget of the active controller to $\hat{\mathcal{B}}_a = 1$ and exploring its optimal allocations for different combinations of the levels of bias b_A and

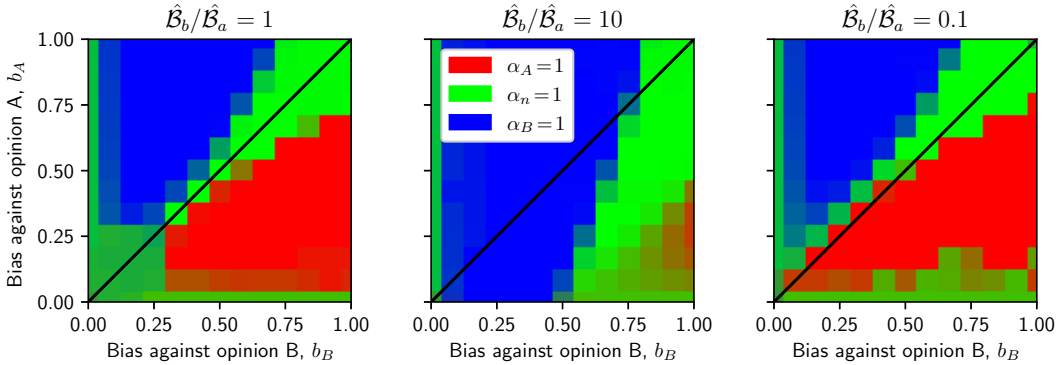


FIGURE 4.6: Target group of optimal allocations by the active controller for different combinations for the levels of bias b_A and b_A, in scenarios of **a** budget equality, **b** budget inferiority, and **c** budget superiority and on a complete graph of size $N = 100$. The passive opponent distributes its budget equally among all agents in the network, i.e. $\beta_A = \rho_A$, $\beta_n = \rho_n$, $\beta_B = \rho_B$.

b_B, in their whole range of possible values and for three scenarios (Fig. 4.6): budget equality ($\hat{\mathcal{B}}_a = \hat{\mathcal{B}}_b$), budget inferiority ($\hat{\mathcal{B}}_b = 10\hat{\mathcal{B}}_a$), and budget superiority ($\hat{\mathcal{B}}_b = 0.1\hat{\mathcal{B}}_a$). This first analysis assumes that the passive opponent distributes its budget evenly across all agents — i.e. $\beta_A = \rho_A, \beta_n = \rho_n, \beta_B = \rho_B$.

When in budget equality (Fig. 4.6**a**), four distinct phases of optimal allocations can be seen. Phase I and Phase II (top**a** and bottom–right of Fig. 4.6**a**, respectively) correspond to an imbalance in the levels of bias between the two biased groups. In either case, the active controller targets the least biased of the two groups ($\alpha_B = 1$ and $\alpha_A = 1$, respectively), as it will be easier to control. Phase III covers the regime of similar levels of bias among the two biased groups (top–right of Fig. 4.6**a** and most of the diagonal) and is characterised by a full targeting of normal agents by the active controller (i.e. $\alpha_n = 1$). When both levels of bias are very high (top–right of the corner), biased agents are very hard to control and all influence is allocated on normal agents. Last, Phase IV spans the scenario where both levels of bias are low and is characterised by the active controller splitting its influence evenly among the three groups, mimicking the opponent's strategy (lower–left corner of the Fig. 4.6**a**).

When in budget inferiority (Fig. 4.6**b**), the phase diagram is qualitatively similar to that of budget equality, but with an expansion of Phase I, entailing a strong preference towards targeting agents biased against the opponent, even when their bias is higher than the bias against the active controller (blue area below the diagonal). Phase II almost vanishes completely, with only a slight preference to agents moderately biased against the active controller ($b_A < 0.5$) if the bias against the opponent is very large ($b_B \approx 1$). Phase IV of equal targeting all agents completely vanishes. So, in short, when in budget inferiority, a strong preference towards targetting agents biased against the opponent is prevalent.

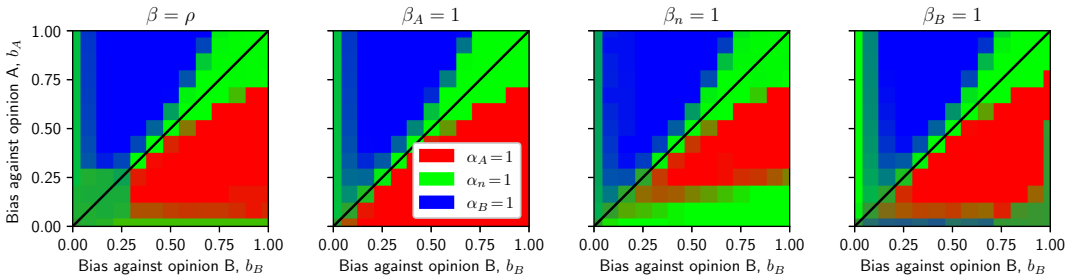


FIGURE 4.7: Target group of optimal allocations by the active controller for different combinations for the levels of bias b_A and b_A, when the **a** opponent targets all agents equally ($\beta = \rho$) or focuses all her influence on **b** agents biased against the active controller ($\beta_A = 1$), **c** normal agents ($\beta_n = 1$), or **d** agents biased against her opinion ($\beta_B = 1$), on a complete graph of size $N = 100$ and budgets $\hat{\mathcal{B}}_a = \hat{\mathcal{B}}_b = 1$.

When in budget superiority (Fig. 4.6**b**), the phase diagram is very similar to when in budget equality, with the exception that i) Phase II is slightly shifted above the diagonal — reflecting a preference towards agents biased against the active controller when both levels of bias are similar — and ii) Phase IV vanishes completely — rejecting the strategy of mimicking the passive opponent by equal targeting.

We then explore variations in the opponent strategy in the scenarios of budget equality and budget inferiority[1] by allowing her to concentrate all her influence in one of the three groups. When in budget equality, there is little variation in optimal strategies if the opponent changes her strategy (Fig. 4.7). We note that Phase IV of targetting all agents equally is only present when the opponent also does so (Fig. 4.7**a**), while a new Phase V of targeting normal agents appears for low levels of b_A and when the opponent targets normal agents (low green band of Fig. 4.7**c**), which reflects a shadowing behaviour when the situation is favourable for the controller (strong imbalance of bias of the population against the opponent). Likewise, a Phase V of targeting normal agents also appears for low levels of b_B and when the opponent targets agents biased against herself (left–hand green band of Fig. 4.7**d**). This reflects an avoidance behaviour of the opponent's targets in an unfavourable situation (strong imbalance of bias of the population against the active controller).

When in budget inferiority, the changes in optimal strategies are qualitatively similar but more prominent (Fig. 4.8). When the levels of bias against the active controller are low (lower side of the phase diagrams), a tendency to shadow the opponent are clearly visible when the opponent targets agents biased against the active controller (Fig. 4.8**b**, red lower–right corner) or when she targets normal agents (Fig. 4.8**c**, green lower–right corner). Additionally, Phase V corresponding to an avoidance behaviour is much wider when the opponent targets agents biased against herself (left–hand band of the Fig. 4.7**d**), as these agents are not the easiest to control any more since they are also targeted by a strong opponent.

[1]Due to the strong similarities of the scenario of budget superiority to that of budget equality, we relegate the results of the earlier to Appendix B.1

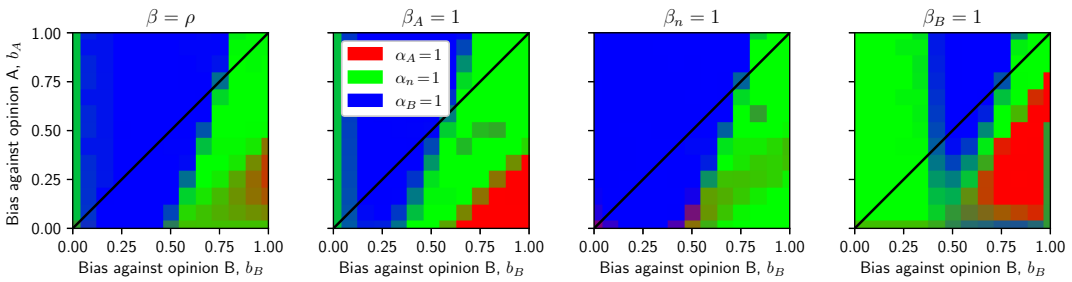


FIGURE 4.8: Target group of optimal allocations by the active controller when in budget inferiority ($\hat{\mathcal{B}}_a = 0.1\hat{\mathcal{B}}_b$), for different combinations for the levels of bias b_A and b_A, when the **a** opponent targets all agents equally ($\beta = \rho$) or focuses all her influence on **b** agents biased against the active controller ($\beta_A = 1$), **c** normal agents ($\beta_n = 1$), or **d** agents biased against her opinion ($\beta_B = 1$), on a complete graph of size $N = 100$.

In summary, when two controllers — one active and one passive — are present, the active controller tends to target the agents that are easiest to control when her situation is unfavourable and targets agents that are harder to control when in a favourable situation. Which agents are easier or harder to control depends on the interplay between level of bias of the agents and the opponent's behaviour; for example, agents that are biased against the opponent are the easiest to control unless their level of bias is low and the opponent targets them — or even if their level of bias is moderate if the opponent has a much larger influence power than the active controller. Hence, we can also see here a similar shadowing–avoidance behaviour as explored in Chapter 3 that interplays with the levels of bias.

4.8 Discussion

In this chapter, we have explored influence maximisation on heterogeneous populations where some agents are biased against one of the opinions. We have first studied a simpler scenario with a single external controller who aims to control a population with some agents biased against the controller's opinion. For this scenario, we have derived an analytical dependency of the optimal strategy on the other parameters on complete graphs, to then explore in more detail the role of node degree on bipartite networks, and degree heterogeneity in Barabasi–Albert networks. Last, we have studied a more complex scenario where groups of agents biased against either opinion co-exist, while another passive external controller who opposes the active controller is also present, and studied optimal allocations under all possible variations of this richer case.

Based on all these experiments, a general pattern in optimal influence allocations can be noted. We find that agents that are harder to control — agents biased against the active controller, hubs in a heterogeneous network, or agents targeted by a strong opponent — receive more influence allocation when the controller is in a favourable

position (due to a large budget or a population more biased against the opponent's opinion), while agents that are easier to control — unbiased agents, nodes in the periphery, or agents biased against the opponent's opinion — receive more allocation when the controller is in an unfavourable position (small budget or a population more biased against the controller's opinion). The transition points between both regimes generally depend on the levels of bias, the budgets of the controllers, and the fraction of biased agents in the network.

Our findings fit in the general picture shown in the previous chapter, where shielding and shadowing also reflect a similar hierarchy of 'difficulty in targeting' and where the order of preference in the hierarchy gets reversed if the controller's availability is low and she is in a disadvantageous situation. Analogous observations have also been seen in previous literature from (Brede et al., 2018) and Brede et al. (2019a), which found that optimal allocations tend to depend on a trade–off between budget availability and the difficulty to control agents.

Although we have uncovered the essential effects that different levels of bias have on influence maximisation, we have limited ourselves to a set of simple scenarios to decouple the role that the different ingredients have in optimal allocations and that can serve as baselines for comparison. As we have developed this basic understanding, more complex scenarios combining all the ingredients here explored (heterogeneous network topologies, agents biased to either opinion, two external controllers) can be explored next, with an aim to better reflect real scenarios and link the model to specific real–world situations. Further extensions to enrich the model could include a larger heterogeneity in agents' bias, the possibility of agents to be 'biased against both opinions simultaneously' (i.e. being stubborn and less susceptible to social influence in opinion change) or including internal opinions (or 'innate preferences').

Chapter 5

The effects of party competition on consensus formation

An abridged version of this chapter has been accepted as a workshop paper to the *3rd International Workshop on Agent–Based Modelling of Human Behaviour (ABMHuB'2021)*, as part of the *Alife 2021* conference (Romero Moreno et al., 2021b).

5.1 Introduction

Opinion polarisation has been a major concern in modern democracies over the last years, and is even considered a threat to their stability by some authors (Abramowitz and Saunders, 2008; Hare and Poole, 2014; Ramos et al., 2015a; Carothers and O'Donohue, 2019). Although moderate levels of polarisation have been positively associated with some indicators of democratic performance such as turnout (Wilford, 2017), too much polarisation can lead to social fragmentation, political division, and the inability to reach any compromise between the parties (Somer and McCoy, 2018). In countries such as the US, recent studies show that polarisation has become a major force in shaping voters' attitudes towards policy issues, parties, and even other voters (Campbell, 2018; Svolik, 2019; Iyengar et al., 2019).

In this context, researchers are interested in determining the role that political actors (Bischof and Wagner, 2019; Hegselmann and Krause, 2015), social media (Sikder et al., 2020; Del Vicario et al., 2017), or external agents (Bhat and Redner, 2019; Gaitonde et al., 2020b) have played in splitting societies' opinions. Evidently, the need for finding the causes (and ascertaining the remedies) of polarisation has caught the attention of many researchers from political science (Layman et al., 2006; Boxell et al., 2017) and sociology (Schweighofer et al., 2020a).

In this chapter, we study the interrelated dynamics of two processes affecting opinion dynamics when multiple issues are debated. First, we consider a model to describe the dynamics of citizens' opinion in the context of consensus formation. Among the models and micro–foundations employed in the opinion dynamics field, the *bounded–confidence* (BC) model has a long tradition in the study of consensus and polarisation (Ramos et al., 2015a; Deffuant et al., 2002). This interaction model includes a *tolerance threshold* that determines the maximum distance in opinion for any two citizens to interact. The tolerance threshold is a key parameter in system outcomes, with decreasing values leading population from consensus towards polarisation and fragmentation (Weisbuch et al., 2003).

Second, our other focus in this work is on the effects of party competition via campaigning on voters. In political science, most research on party competition originates from spatial voting theory as introduced by Downs (1957), where citizens hold political positions on a multidimensional political space and vote for parties according to the distance between their opinions and those of the parties. In this context, party competition is mainly investigated from two different approaches. In the first, parties are considered to compete for votes by *moving* towards positions that can attract larger vote shares (Laver and Sergenti, 2011; Miller and Stadler, 1998; Adams and Merril, 1999; Alvarez et al., 2000). However, parties are constrained within their position because changing them might be costly, as they may lose credibility to the electorate (De Sio and Weber, 2014). In the second approach, parties compete for vote shares by *adjusting the saliency* given to the different political dimensions (e.g. by campaigning) to promote aspects of the debate in which they hold favourable positions relative to the electorate (Dragu and Fan, 2016; Amorós and Puy, 2010; Feld et al., 2014; De Sio and Weber, 2014).

In this chapter, we unite two disjoint branches of research by studying how party competition interferes with processes of consensus–formation in a two–dimensional political opinion space. More specifically, we investigate a model of party competition via saliency adjusting and couple its dynamics to a model of consensus–formation and polarisation via the bounded–confidence model. This setting allows us to explore the effect of external controllers in opinion dynamics systems from the perspective of adaptive controllers that react to the opinion dynamics in similar time scales. We discover that the effects of party competition on consensus formation may be very different depending on party positions in the political space. Below, we illustrate that party competition can foster consensus formation in some scenarios, but may also promote polarisation in others. We find that these differences in effects are strongly linked to the adaptive behaviour of saliency changing by the parties, with shifts in saliency promotion strategies favouring consensus. We further explore the characteristics of the constellation of party positions that lead to either behaviour, and

expand the inspect the role of the size of the party system in the outcomes. Last, we explore how embedding the agents into a sparse social network affects the dynamics.

5.2 A Model for party competition and consensus–formation processes

Our purpose is to model interactions between party and opinion dynamics in the formation of consensus within a society. First, we consider a population of citizens or voters that hold individual political opinions, modelled as multidimensional continuous opinions, an assumption common in both the opinion dynamics (Castellano et al., 2009a) and political science (De Sio and Weber, 2014) fields. Examples of typical political dimensions regarded in the political science literature are the *economic left–right*, or the *social GAL-TAN* (Green–Alternative–Liberal, Traditional–Authoritarian–Nationalism). We then let opinions in these dimensions evolve subject to peer interactions following the well–known bounded–confidence model (Weisbuch et al., 2003).

Second, we also consider a set of parties, each of which holds a fixed opinion in the political space. Parties compete for votes by adjusting the emphasis with which they promote the various political dimensions. In other words, they affect the perceived importance —or *saliency*— of the dimensions. The role of saliency competition has been emphasised by many political scientists in the past (Amorós and Puy, 2010; De Sio and Weber, 2014; Feld et al., 2014). We assume that citizens cast their vote in a probabilistic fashion according to their perceived distance to the parties (Burden, 1997; Alvarez and Nagler, 1998). Parties then promote the different dimensions as to increase their vote share by promoting dimensions in which they have a favourable position with respect to the electorate, thus changing the perception of distances in the population. Hence, the dynamics of party competition and consensus formation interact, as the consensus dynamics is affected by perceived distances between citizens' positions. We proceed by giving details and a formalisation of the models of opinion and party competition below.

5.2.1 Opinion dynamics model

Consider a population of citizens whose opinions are initially uniformly spread in a D-dimensional space of opinions constrained to the interval $[-1, 1]$, i.e. $\boldsymbol{o} \in O = [-1, 1]^D$. Opinion dynamics follow the *bounded–confidence* (BC) model introduced by Deffuant et al. (2000), although extended to multiple dimensions. This implies that a distance metric $\mathrm{d} : O \times O \to \mathcal{R}$ needs to be defined and the update of

opinions when two randomly chosen agents i and j encounter at each time step t is given by

$$\boldsymbol{o}_{i,t+1} = \begin{cases} (1-\alpha)\,\boldsymbol{o}_{i,t} + \alpha\,\boldsymbol{o}_{j,t} & \text{if } \mathrm{d}(\boldsymbol{o}_{i,t}, \boldsymbol{o}_{j,t}) \leq \delta \\ \boldsymbol{o}_{i,t} & \text{otherwise,} \end{cases} \tag{5.1}$$

where $\alpha \in [0, 1/2]$ is the concession parameter and citizen j also simultaneously updates her opinion $o_{j,t+1}$ in an analogous move. We assume that distances in opinion are affected by the perception of the importance of the various dimensions, represented by a normalised vector of saliencies $w_d \in [0,1], d = 1, ..., D$ that sums up to one. Accordingly, we employ a normalised weighted Euclidean distance[1] as our distance metric, given by

$$\mathrm{d}(\boldsymbol{o}_i, \boldsymbol{o}_j) = \frac{1}{2}\sqrt{\sum_{d=1}^{D} w_d \left(o_i^{(d)} - o_j^{(d)}\right)^2}\,. \tag{5.2}$$

The basic formulation of the BC model assumes that any two citizens can encounter. A more realistic variation, would restrict the possibility of encounters by defining a social network where agents are identified with nodes and edges define possible pairwise encounters. Therefore, at each time step t, a random edge is selected and the citizens that are linked by it have an encounter. More details about the BC model, see Sect. 2.2.2.

5.2.2 Model of party competition

On top of the population of voting citizens, we also consider a number of parties $k = 1, \ldots, K$ that have fixed positions in the opinion space $\boldsymbol{p}_k \in O$. According to probabilistic models of spatial voting (Burden, 1997; Alvarez and Nagler, 1998), party k is assumed to receive a vote from citizen i at time t, i.e. $v_{i,t} = k$, with a probability that is related to the distance of their respective opinions and the distance of the citizen's opinion to the other parties:

$$P(v_{i,t} = k) = \frac{\exp(-\gamma\,\mathrm{d}(\boldsymbol{o}_{i,t}, \boldsymbol{p}_k))}{\sum_{l=1}^{K} \exp(-\gamma\,\mathrm{d}(\boldsymbol{o}_{i,t}, \boldsymbol{p}_l))}\,, \tag{5.3}$$

where γ parametrises the range of party influence. For large γ, parties have a small range of attraction around their core positions in opinion space; for small γ, parties can attract voters with more strongly deviating opinions[2].

[1] Fortunato et al. (2005) compared results between a 1-norm Manhattan and a 2-norm Euclidean distance metrics on a 2D Hegselmann-Krause bounded–confidence model and found little effect of this decision on the resulting stationary clusters.

[2] Note that with $\gamma \to \infty$, a deterministic model of voting is recovered, where votes are always given to the party whose position is closest to the citizen. Lower values of γ imply the notion of uncertainty in choice, as there may be hidden attributes not modelled in voting space or voters may not vote for "what

We assume that each party k has a vector of saliency promotion $\boldsymbol{w}^{(k)} = (w_1^{(k)}, \ldots, w_D^{(k)})$ normalised to $\sum_{d=1}^{D} w_d^{(k)} = 1$, as to model a resource constraint in party campaigning. Following Amorós and Puy (2010) and Dragu and Fan (2016), citizens aggregate the overall effect of all parties' campaigning efforts —which determines the overall saliency $\boldsymbol{w}$ in the population— via a function $\boldsymbol{f} : [0,1]^{D\times K} \to [0,1]^D$, which is assumed to be differentiable and monotonic. For the purposes of this study, we consider the simple aggregation function $w_d = 1/K \sum_{k=1}^{K} w_d^{(k)}$, where the total saliency of each dimension w_d is the average of saliency promotions given by parties to that dimension.

Each party k attempts to maximise its expected vote share from the electorate at each time step t by directing attention to political dimensions beneficial to gaining votes, i.e. by adjusting its vector of saliency promotion $\boldsymbol{w}_t^{(k)}$ through campaigning. In the following, we consider parties to be myopic; i.e. we assume that they attempt to maximise their expected vote share at the current time $V_{k,t} = 1/N \sum_{i=1}^{N} P(v_{i,t} = k)$ without considering the vote share at a future time $t' > t$ and not accounting for the strategy of the opponents or the dynamics of opinions. This reflects parties with bounded rationality and limited information (Laver and Schilperoord, 2007). We assume that parties adjust their saliency promotion at a time scale τ, modelled by implementing changes in saliency promotion only after every τ time steps of the opinion dynamics.

To model rational parties that have local knowledge of the effects of changing their campaigning efforts, we assume that parties follow a gradient dynamics, adjusting their saliency promotion in the direction of largest gain in expected vote share $\nabla_{\boldsymbol{w}^{(k)}} V_{k,t}$, which is given by

$$\nabla_{\boldsymbol{w}^{(k)}} V_{k,t} = 1/N \sum_{i=1}^{N} \nabla_{\boldsymbol{w}^{(k)}} P_{i,k}^{(t)} = \gamma/N \sum_{i=1}^{N} P_{i,k}^{(t)} \left(\sum_{l=1}^{K} P_{i,l}^{(t)} \nabla_{\boldsymbol{w}^{(k)}} \mathrm{d}_{i,l}^{(t)} - \nabla_{\boldsymbol{w}^{(k)}} \mathrm{d}_{i,k}^{(t)} \right) , \quad (5.4)$$

where $P_{i,k}^{(t)} = P(v_{i,t} = k)$, $\mathrm{d}_{i,k}^{(t)} = \mathrm{d}(\boldsymbol{o}_{i,t}, \boldsymbol{p}_k)$, and $l = 1, \ldots, K$ is a summation index over parties. For the aggregation function introduced above, we have

$$\frac{\partial\, \mathrm{d}_{i,l}}{\partial\, w_d^{(k)}} = \frac{1}{8\,K} \left(o_i^{(d)} - p_l^{(d)} \right)^2 \mathrm{d}_{i,l}^{-1} . \quad (5.5)$$

Specifically, after every τ updates of the bounded–confidence dynamics of opinion formation, parties adjust saliencies via

$$\boldsymbol{w}_{t+1}^{(k)} = \boldsymbol{w}_t^{(k)} + \epsilon/\gamma\, \nabla_{\boldsymbol{w}^{(k)}} V_{k,t} , \quad (5.6)$$

is best for them", i.e. they may not be fully rational. Similar forms to (5.3) are common in choice theory (Anderson et al., 1992; McFadden, 1994)

which is then projected[3] into the budget constraint $\sum_{d=1}^{D} w_d^{(k)} = 1$ and the feasible region $w_d^{(k)} \in [0, 1]$. In the above, we introduce the parameter $\epsilon \in \mathbb{R}_{0+}$, which models the time scale at which parties can perform saliency adjustments. Note that a choice of $\epsilon = 0$ retrieves the BC model in which opinions evolves without any effect from party competition[4].

5.3 Methods and experimental settings

We explore the proposed model and its outcomes via agent–based simulations, as analytical solutions are hard to obtain due to the non-linearities and complex interactions of the model. We will perform experiments on the case of $D = 2$ dimensions of the opinion space. This setting already serves to illustrate the richness of possible model outcomes while still allowing intuitive understanding of the model behaviour via visual inspection of the evolving opinions. Moreover, two–dimensional opinion spaces are pervasive in political science research, since the standard assumption is that political competition is spanned over an economic and a socio–cultural dimension in most democratic countries (Enyedi and Deegan-Krause, 2010). We will perform most of our analysis with $K = 3$ political parties, which is a typical configuration — e.g. the political system in the UK or in Spain before the 2008 crisis. Regarding the size of the population, although we are interested in political systems potentially comprising millions of citizens, we set their number to $N = 10^5$, as a trade–off between computational constraints and the quality of the results. A set of experiment testing other system sizes can be found in Appendix C.1.

The parameters α, γ, τ, and ϵ relate to the time scales at which both types of dynamics — citizens' opinions and party saliency promotions — evolve, while they do not affect other aspects of the dynamics. As we are interested in the interaction between both types of dynamic, we want them to occur at similar time scales and we select these four time–scale parameters in a trial–and–error process that allow for this interaction while keeping simulation times manageable. These are a concession rate at $\alpha = 1/4$, the range of party influence in voting preferences at $\gamma = 10$, the saliency–changing period at $\tau = N/2$, and the speed of saliency change at $\epsilon = 64$.

Regarding initial conditions at the start of each simulation, we sample initial opinions of citizens from a uniform probability distribution over the opinion space O. Party opinions are also sampled this way, although they remain fixed throughout the simulation, and they start promoting all political dimensions with equal intensity, $w_{1,t=0}^{(k)} = w_{2,t=0}^{(k)} = 1/2$ (i.e. a neutral saliency mix). We run simulations until the

[3] The projection is implemented by setting negative values to zero and normalising the resulting vector.

[4] Note that we have introduced two parameters (ϵ and τ) related to the time scale at which parties adjust their promotion strategies. The purpose of such redundancy of parameters is to limit computations of gradients and thus improve computational efficiency.

stationary state has been reached, which we define by two conditions. First, the difference in the standard deviation of citizens' opinion in the last 50τ time steps must be lower than 10^{-4}, i.e. $[\text{std}(\boldsymbol{o}_{t-50\tau}) - \text{std}(\boldsymbol{o}_t)] < 10^{-4}$. Second, total absolute changes in saliency promotion by each party in the last 50τ time steps must be lower than 0.005, i.e. $\sum_d |w^{(k)}_{d,t-50\tau} - w^{(k)}_{d,t}| < 0.005$.

Since the stochastic elements of the model (e.g. choosing two random agents at each time step) may affect the outcomes, it is important that enough simulations and statistical analyses are performed to provide enough confidence that what is observed is not a result of chance alone. Unless otherwise specified, for each parameter combination in each experiment, we run 30 simulations and show averages of the outcome variables, along with standard deviations.

Regarding the outcomes of the model, we are mainly interested in the phenomena of populations arriving to a state of consensus, polarisation, or fragmentation. Following previous work on the BC model (Lorenz, 2007), these outcomes can be operationalised by analysing the number of clusters in the stationary state of the simulation. However, defining clusters is not a univocal task. Here, we exploit the role of the confidence threshold δ in preventing further interaction and identify clusters by dividing the space into a $\lfloor 2/\delta \rfloor \times \lfloor 2/\delta \rfloor$ grid, thus ensuring that the maximum separation of any two opinions within a cell of the grid is approximately $\sqrt{2}$ times smaller than δ (in the two–dimensional case) and hence any two citizens whose opinions are within the same grid can still potentially interact. We define a cluster as groups of consecutive cells in this grid, while different clusters are separated by a band empty cells in the grid. We ignore fine–grained noise that remains due to numerical methods by regarding cells that contain less than 0.5% of the total population as empty.

Still, the number of clusters alone may be misleading, as it can be influenced by the occurrence small clusters, so we also incorporate the information of clusters size to define the *effective number of clusters*, given by $\phi = 1/\sum_c (s_c/N)^2$, where s_c is the size of cluster c. With this measure, full consensus corresponds to $\phi \sim 1$, and paradigmatic (maximum) polarisation to $\phi \sim 2$, meaning that the population is divided into two clusters of similar size. Values of $\phi \in (1, 2)$ generally entail the presence of two clusters of unequal size, and $\phi > 2$ corresponds to fragmentation into more than two clusters. In this chapter, we will consider an outcome of 'consensus' for $\phi \leq 1.5$, an outcome of 'polarisation' for $1.5 < \phi \leq 2.5$ and an outcome of 'fragmentation' for $\phi > 2.5$.

In previous research on the bounded–confidence model, it has been shown that the tolerance threshold δ is a key parameter whose values determine the regimes of these three outcomes (Hegselmann and Krause, 2002; Weisbuch et al., 2003; Lorenz, 2007). Therefore, we define the transition value $\delta_{\phi:1\to 2}$ as the middle point between the lowest δ that results in consensus ($\phi \leq 1.5$) and the highest δ that results in polarisation ($\phi > 1.5$). We define an analogous transition value $\delta_{\phi:2\to 2+}$ as the middle

point between the lowest δ that results in polarisation ($\phi \leq 2.5$) and the highest δ that results in fragmentation ($\phi > 2.5$).

In our experiments here, we aim to explore how the inclusion of party dynamics affects the outcomes of the opinion dynamics. For this purpose, we first generate many random party configurations and characterise the different alterations that party dynamics have on opinion outcomes (Sect. 5.4), finding three possible effects: fostering polarisation, fostering consensus, or not altering the opinion dynamics. To better understand the mechanisms behind these effects, we analyse in more detail single runs of each, thus gaining insight of of the interactions between the saliency strategies and opinion dynamics (Sect. 5.5). A behaviour that we hypothesise is behind the fostering of consensus is a change of in the saliency promotion of parties during the simulation. We thus measure the *number of saliency changes* σ_k of party k as the number of times two consecutive updates in its saliency promotion have different signs, i.e. $\sigma_k = \frac{1}{2}\sum_{m=1}\left(1 - sign\left(w^{(k)}_{d=1,t=m\,\tau}\right) sign\left(w^{(k)}_{d=1,t=(m-1)\,\tau}\right)\right)$.

In Sect. 5.6, we go one step further by exploring which characteristics of the party configurations may be behind producing the different saliency–changing behavours. To test for different characteristics of the party systems, we curated a set of features related to the location and geometry of the party system, including the following:

- The Euclidean distance of each party to the centre of the space and their mean distance to the centre of the space (4 features).
- The diagonality of each party opinion, i.e. whether they are closer to the principal axes (zero diagonality) or to the diagonals (diagonality equal to one), and the mean diagonality of the party system (8 features).
- Angles of the triangle that the three party opinions form (3 features).
- Area of the triangle that the three party opinions form (1 feature).
- Pairwise distances between the three party opinions (3 features).
- Pairwise absolute differences between party angles when polar coordinates are used (3 features).
- Distance to the closest boundary of each party (3 features).
- Distance of the centre of mass of the three parties to the centre of space (1 feature).

Using the above features, we train a decision tree (Krzywinski and Altman, 2017) to classify party configurations into the above three possible effects. We use a decision tree due to its high interpretability and its ability to automatically find the most divisive features. We use the entropy as a measure the quality of splits and, to have a

balanced dataset, we give weights to each sample inversely proportional to the number of samples in its corresponding class. To keep a high interpretability of the tree and prevent over–fitting, we restrict the depth of the tree to a maximum level of 4 and only keep leaves if they contain more than 10% of the (weighted) samples.

In Sect. 5.7, we extend experiments to other numbers of parties K to examine how the size of the party system affects their influence in the opinion dynamics, by performing a similar analysis as in Sect. 5.4 and contrasting results among the different K.

Our experiments assume that citizens can communicate with any other citizen indistinctly. This assumption is particularly unrealistic when considering large populations, so we finally perform a set of experiments where we impose a social network structure onto the population — with interactions only possible between connected agents — and explore how this modification alters previous results. In these, we build Erdos–Renyi random networks, which are commonly used in agent–based models of opinion dynamics (Amblard et al., 2015), with a probability $\rho = \langle d \rangle / N$ of each link to exist as to produce an expected node degree of $\langle d \rangle = 10$, i.e. $\rho = 10^{-4}$.

We conclude the section with a summary of all parameters of the model and the values chosen for our experiments (Table 5.1). The code used to run all experiments and produce the figures can be found at
`https://git.soton.ac.uk/grm1g17/bc-polarisation`.

Parameter	Definition	Values in our experiments
$D \in \mathbb{N}$	Number of dimensions	$D = 2$
$K \in \mathbb{N}$	Number of parties	$K = 3$ (except Sect. 5.7)
$N \in \mathbb{N}$	Number of citizens	$N = 10^5$
$\delta \in [0, 1]$	BC tolerance threshold	$\delta \in [0.12, 0.245]$ (varying)
$\alpha \in [0, 1/2]$	Concession rate upon interaction	$\alpha = 1/4$
$\gamma \in \mathbb{R}_+$	Range of party influence	$\gamma = 10$
$\tau \in \mathbb{N}$	Strategy–changing period	$\tau = N/2$
$\epsilon \in \mathbb{R}_{0+}$	Change rate of saliency promotion	$\epsilon = 64$
$\rho \in [0, 1]$	Probability of a link to exist	$\rho = \langle d \rangle / N = 10^{-4}$

TABLE 5.1: Parameters of the model and the values we employ in our experiments (if not specified otherwise when introducing an experiment). The last parameter, ρ, only affects to the last set of experiments, which include a social network structure.

5.4 Effects of party competition on consensus formation

We begin our experiments with an analysis of how the presence of party dynamics affect the outcomes of consensus, polarisation, and fragmentation. For this purpose, we generate 500 random party configurations and characterise the alterations that party dynamics make to the transition values $\delta_{\phi:1\to2}$ and $\delta_{\phi:2\to2+}$. For each party

configuration, we run a grid of experiments in the range $\delta = [0.155, 0.235]$ in intervals of $\Delta\delta = 0.05$, with 10 simulations per δ and use these to find $\delta_{\phi:1\to2}$ and $\delta_{\phi:2\to2+}$ as the middle points between consecutive δs that produce different outcomes.

Figure 5.1**a** shows results for these experiments including the combinations of transition values for all the sampled party configurations. We analyse first results regarding the value of δ that marks the transition from consensus to polarisation ($\delta_{\phi:1\to2}$), which corresponds to the x-axis of the plot. As a reference, the case with no party dynamics is shown by a black cross, with values $\delta_{\phi:1\to2} = 0.2025$ and $\delta_{\phi:2\to2+} = 0.1975$. Cells to the left of the cross, with $\delta_{\phi:1\to2} < 0.2$, imply that the party configuration led to an outcome of consensus for values of δ where not having a party competition would have led to polarisation, i.e. the party dynamics are 'fostering consensus'. These comprise 86 (17.2%) of the 500 sampled party configurations. In contrast, cells to the right of the cross, with $\delta_{\phi:1\to2} > 0.205$, contain party configurations that led to a an outcome of polarisation for values of δ where not having party competition would have led to consensus, i.e. the party dynamics are 'fostering polarisation'. These comprise 351 (70.2%) of the 500 sampled party configurations. Furthermore, the outcome of $\delta_{\phi:1\to2} = 0.2275$ and $\delta_{\phi:2\to2+} = 0.1675$ is particularly common, with 180 (36%) party configurations producing it (yellow square on the right side of Fig. 5.1**a**). Cells in the middle band, with $\delta_{\phi:1\to2} = 0.2025$, are party dynamics that did not modify the transition value $\delta_{\phi:1\to2}$, comprising 63 (12.6%) cases, out of which 25 (5%) did not alter neither $\delta_{\phi:1\to2}$ nor $\delta_{\phi:2\to2+}$ as compared to the baseline of no party competition.

Regarding the values of δ that mark the transition from polarisation to fragmentation ($\delta_{\phi:2\to2+}$, y-axis), we note that party competition never led to an outcome of fragmentation where its absence would have led to polarisation (or consensus), which can be seen in the absence of cells above the 'no-party' cross. Moving down from the 'no-party' cross implies that party configurations led to regimes of polarisation where no party dynamics would have led to fragmentation.

Next, we will examine in more detail the mechanisms that result in party competition fostering of polarisation or consensus.

5.5 The mechanisms behind the fostering–polarisation or fostering–consensus effects of party competition

To examine in more detail the mechanisms behind fostering polarisation or consensus, we will analyse two party configurations that lead to either outcome, plus a party configuration that does not modify the opinion dynamics. To examine the fostering of polarisation, we select one of the party configurations $\boldsymbol{p}^P$ from the group with the

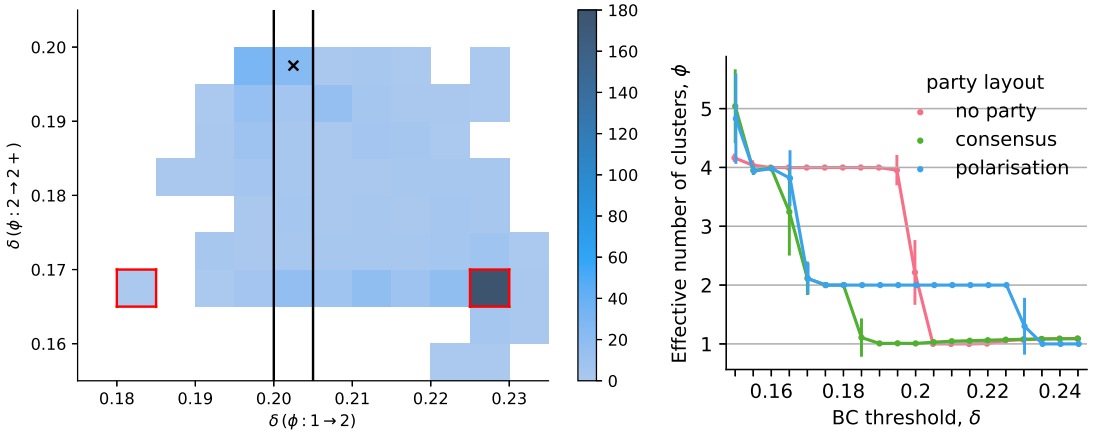


FIGURE 5.1: **a** Histogram of phase transitions $\delta_{\phi:1\to2}$ and $\delta_{\phi:2\to2+}$ for 500 randomly sampled party configurations. The black cross indicates the phase transitions of the no–party baseline, while solid lines separate the different classes of outcomes, from left to right: fostering consensus, no effect on the opinion dynamics, and fostering polarisation. **b** Dependence of the effective number of clusters ϕ on the BC tolerance threshold δ for parties positioned in a polarisation–fostering configuration $\boldsymbol{p}^P$ (orange), a consensus–fostering configuration pC (red), or without party competition (purple). Error bars represent the standard deviation over 30 simulation runs.

most common outcome of $\delta^P_{\phi:1\to2} = 0.2275, \delta^P_{\phi:2\to2+} = 0.1675$ (red square on the right side of Fig. 5.1**a**)[5]. To examine the fostering of consensus, we select a party configuration $\boldsymbol{p}^C$ with the strongest fostering–consensus effect ($\delta^C_{\phi:1\to2} = 0.1825, \delta^C_{\phi:2\to2+} = 0.1675$, red square on the left side of Fig. 5.1**a**) [6]. As a reference, we select a party configuration $\boldsymbol{p}^N$ from the group that did not affect the opinion dynamics (red square around the black cross in Fig. 5.1**a**)[7].

The dependence of their resulting effective number of clusters ϕ on tolerance parameters δ for these three party configurations are shown in more detail in Fig. 5.1**b**. There, we can see three interesting regimes with differences to the baseline of no competition: i) party competition resulting in polarisation where its absence would lead to consensus — i.e. the outcome of the fostering–polarisation party configuration $\boldsymbol{p}^P$ in the regime $\delta^* \in [0.205, 0.225]$; ii) party competition resulting in polarisation where its absence would lead to fragmentation — i.e. the outcome of the fostering–polarisation party configuration $\boldsymbol{p}^P$ in the regime $\delta^{**} \in [0.17, 0.2]$; and iii) party competition resulting in consensus where its absence would lead to polarisation or fragmentation — i.e. the outcome of the fostering–consensus party configuration $\boldsymbol{p}^C$ in the regime $\delta^{***} \in [0.185, 0.2]$. We will examine single simulation runs within each of these regimes next to uncover the mechanisms that may be behind these behaviours.

[5]The party opinions of the selected party configuration from $\boldsymbol{p}^P$ correspond to $\boldsymbol{p}_{k=1} = (0.25, 0.79)$, $\boldsymbol{p}_{k=2} = (0.55, -0.55)$, and $\boldsymbol{p}_{k=3} = (-0.4, 0.75)$.

[6]The party opinions of the selected party configuration in $\boldsymbol{p}^C$ correspond to $\boldsymbol{p}_{k=1} = (0.15, 0.34)$, $\boldsymbol{p}_{k=2} = (0.11, -0.58)$, and $\boldsymbol{p}_{k=3} = (-0.64, 0.05)$.

[7]The party opinions of the selected party configuration in $\boldsymbol{p}^N$ correspond to $\boldsymbol{p}_{k=1} = (0.71, -0.89)$, $\boldsymbol{p}_{k=2} = (-0.5, 0.31)$, and $\boldsymbol{p} = N_{k=3} = (0.16, 0.13)$.

The first simulation setting illustrates how party competition can foster polarisation where its absence would result in consensus. In Fig. 5.2, we show in detail the evolution of two simulation runs for $\delta^* = 0.22$ and party configuration pP. The top and middle rows of the figure show three time steps of two analogous simulation runs; one run where there is no party competition (top row) — so the perceived saliency for both dimensions remains equal in the whole simulation — and one run where parties compete to increase their vote share and affect the perceived saliency of topics (middle row). For a fair comparison, we set the same seed of the random number generator for both runs, meaning that initial conditions are identical (c.f. the left panel in both rows), as well as the pattern of citizen pairs chosen to interact at each time step. Panels on the right column illustrate stationary distributions of citizen opinions, with clusters differentiated by colours and whose sizes are shown in the legend[8]. When there is no party competition (top row), we can see that opinions first gather forming a cross with its centre empty (top–middle panel) that then shrinks down into a single cluster i.e. a state of consensus (top–right panel). In contrast, when including party competition (middle row) opinions first gather into two clusters (central panel) that are preserved until the stationary state, arriving at a state of polarisation, with two big clusters of similar size (middle–right panel).

The bottom row of Fig 5.2 gives the evolution of different metrics over time for the scenario with party competition. The evolution of each party's expected vote share is shown in the left panel, the evolution in standard deviation of citizens' opinion in the middle panel, and the evolution of parties' saliency promotion of the first dimension $w_{d=1}$ in the right panel[9]. From the bottom–right panel, it can be seen that, very early in the simulation, the perceived saliency of dimension $d = 1$ (*dashed line*) is higher than $w_d > 0.5$, as parties $k = 2, 3$ give full promotion to that dimension and disregard dimension $d = 2$. Consequently, citizens give more saliency to dimension $d = 1$ also when perceiving their distances to other citizens, leading to two clusters aligned along this dimension (see the interim state from the central panel of the figure). This outcome can be understood if changes in perceived saliency are likened to deformations of the opinion space. When one dimension from the saliency mix is favoured, the opinion space 'elongates' along such dimension and 'shrinks' in the other dimensions, so clusters align along the favoured dimension.

The second simulation setting illustrates how party competition can foster polarisation where its absence would result in fragmentation. Fig. 5.3 shows the evolution of two simulation runs for $\delta^{**} = 0.195$ and party configuration pP, with results laid out in analogy to those in Fig. 5.2 and using the same seed of the random number generator as well. In the top row, we again see the evolution of the model

[8]Note that cluster sizes do not sum up to N as some agents are scattered into groups smaller than 0.5% of the population and therefore not counted as clusters.

[9]Note that parties' promotion of the second dimension $w_{d=2}$ is redundant, as $\sum_d w_d = 1$, i.e. $w_2 = 1 - w_1$.

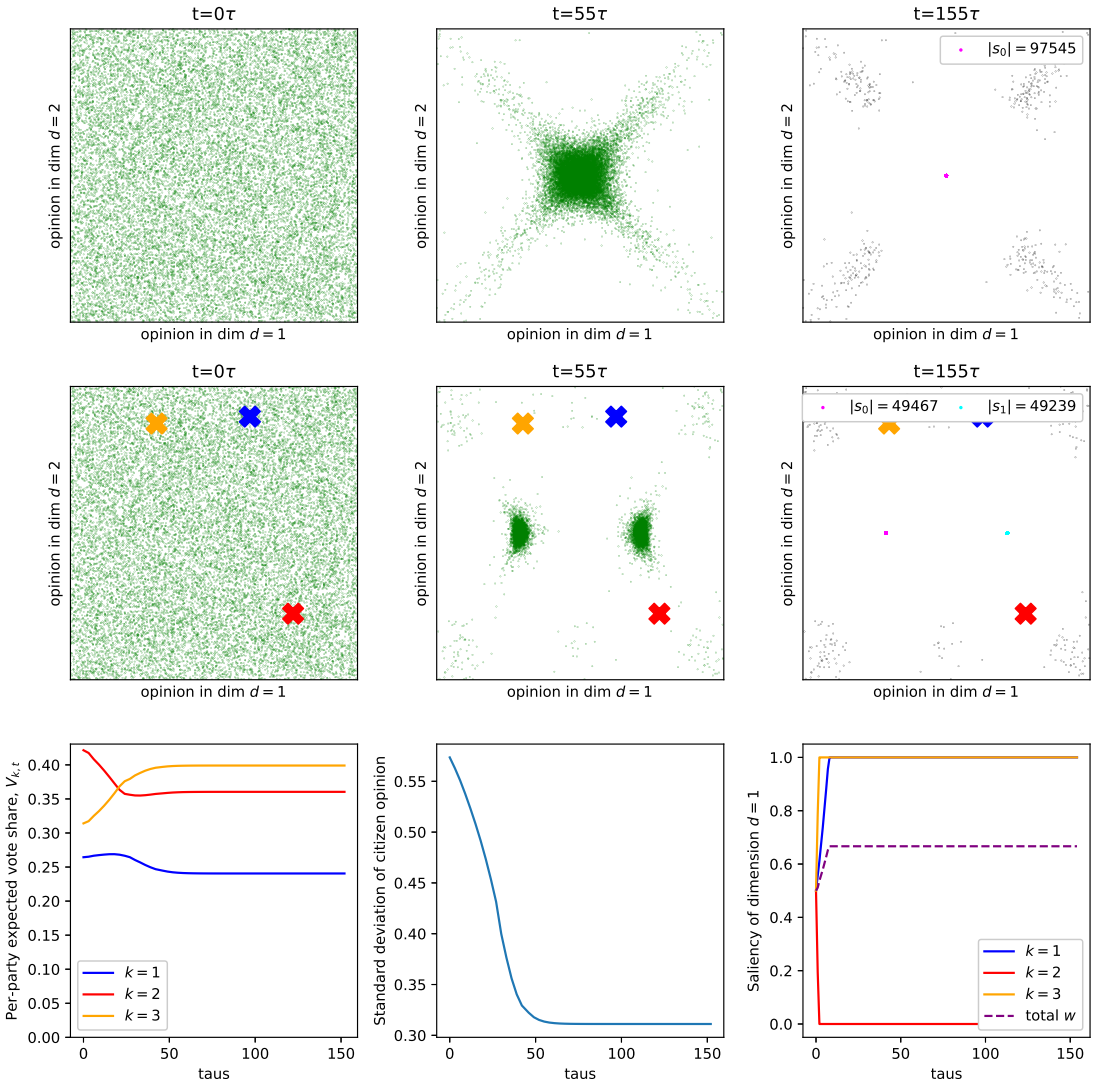


FIGURE 5.2: Simulation run for a fostering–polarisation party configuration ($\boldsymbol{p}^P$) in the regime where party competition fosters polarisation while the population would arrive at consensus in its absence, with $\delta^* = 0.22$. The first two rows show the initial conditions (left column), a middle snapshot of the simulation (middle column), and the stationary outcome (right column) for runs without (top row) and with (middle row) party competition. Coloured crosses indicate parties' opinions and the shown area covers the whole opinion space O. In the right column, opinions are coloured by the cluster they belong to, while legends show cluster sizes. The bottom row shows the evolution of each party's expected vote share (left), dispersion in citizens' opinion (middle), and parties' promotion of dimension $d = 1$ (right).

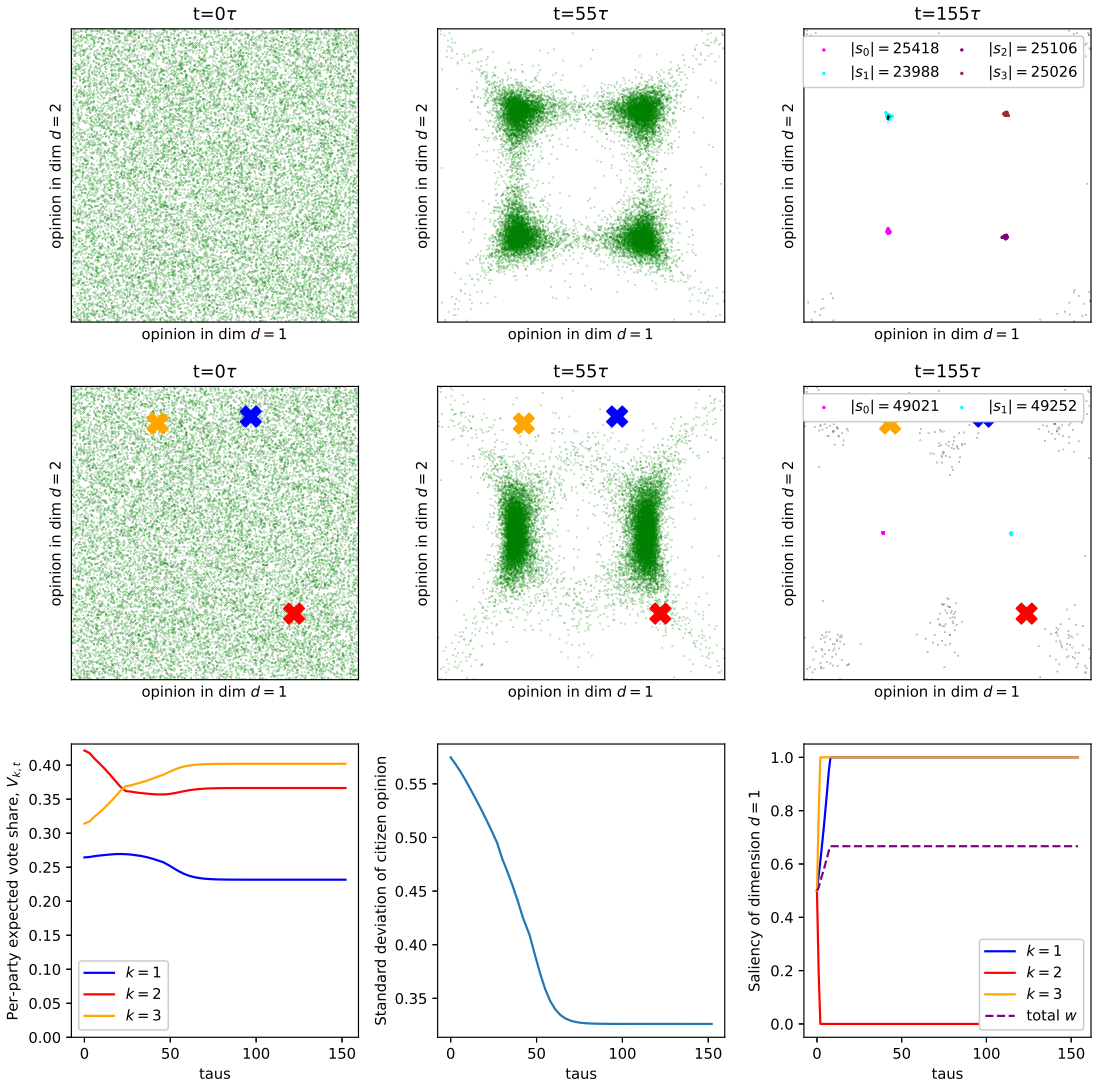


FIGURE 5.3: Simulation run for a fostering–polarisation party configuration ($\boldsymbol{p}^P$) in the regime where party competition fosters polarisation while the population would arrive at fragmentation in its absence, with $\delta^{**} = 0.195$. The first two rows show the initial conditions (left column), a middle snapshot of the simulation (middle column), and the stationary outcome (right column) for runs without (top row) and with (middle row) party competition. Coloured crosses indicate parties' opinions and the shown area covers the whole opinion space O. In the right column, opinions are coloured by the cluster they belong to, while legends show cluster sizes. The bottom row shows the evolution of each party's expected vote share (left), dispersion in citizens' opinion (middle), and parties' promotion of dimension $d = 1$ (right).

without party competition. However, we can see for this smaller tolerance threshold δ^{**} that opinions gather more at the corners of the cross, leaving a wider gap in the centre (top–middle panel). This in turn results in the four corners developing into its own clusters, leading to fragmentation (top–right panel). The inclusion of party configuration $\boldsymbol{p}^P$ leads to a very similar outcome than in Fig. 5.2, arriving at a state of polarisation (middle row). The evolution of the different metrics over time for this setting (bottom row) is also very similar to the case of Fig. 5.2 with $\delta^* = 0.22$.

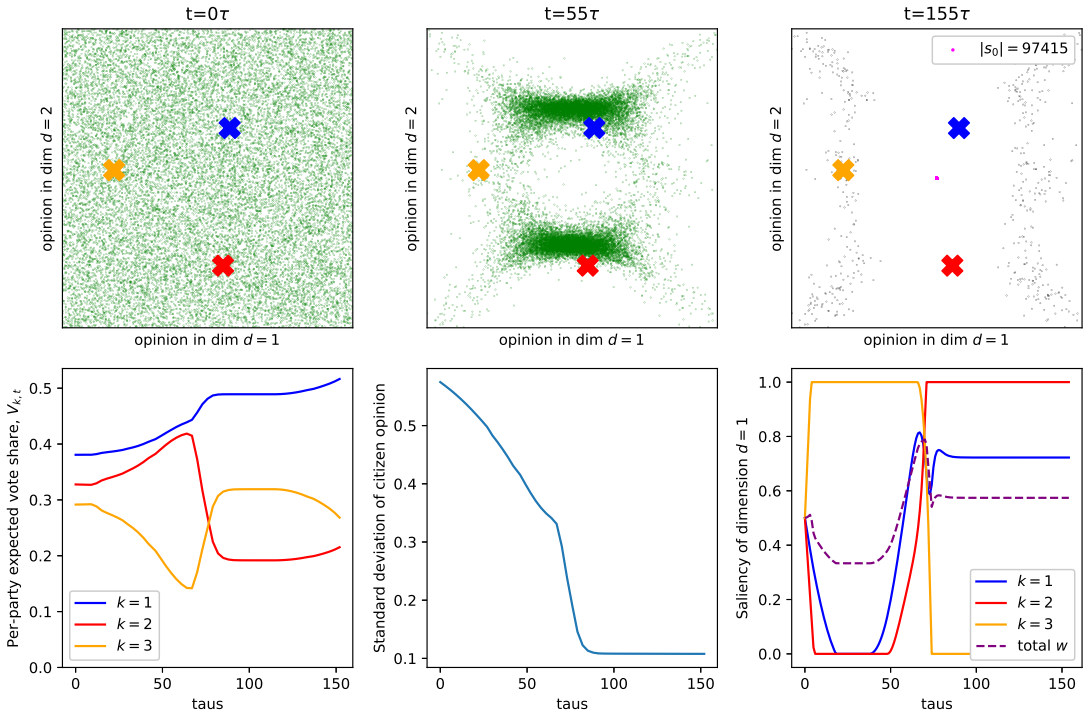


FIGURE 5.4: Simulation run for a fostering–consensus party configuration ($\boldsymbol{p}^C$) in the regime where party competition fosters consensus while the population would arrive at fragmentation in its absence, with $\delta^{***} = 0.195$. The first row shows the initial conditions (left column), a middle snapshot of the simulation (middle column), and the stationary outcome (right column) for runs without (top row) and with (middle row) party competition. Coloured crosses indicate parties' opinions and the shown area covers the whole opinion space O. In the top–right panel, opinions are coloured by the cluster they belong to, while legends show cluster sizes. The bottom row shows the evolution of each party's expected vote share (left), dispersion in citizens' opinion (middle), and parties' promotion of dimension $d = 1$ (right).

The third simulation setting illustrates how party competition can foster consensus where its absence would result in polarisation or fragmentation, as shown in Fig. 5.4, with $\delta^{***} = 0.195$, party configuration pC, and same seed of the random number generator as in Figs. 5.2 and 5.4. As we have selected $\delta^{***} = \delta^{***}$, we omit the row showing the simulation when there is no party competition, which can be seen at the top row of Fig. 5.3. The inclusion of party configuration $\boldsymbol{p}^C$ (middle row) now first drives the population into two clusters (central panel) which then join into a single big cluster corresponding to a state of consensus. The evolution of party saliency in this setting present a characteristic shift in saliency promotion at around $t \sim 40\tau$ (bottom–right panel) and can explain the big difference in effect on the outcomes. Since dimension $d = 1$ was overall favoured before the shift, two clusters along that dimension formed (see the central panel). However, dimension $d = 2$ becomes predominantly promoted after the shift, and most citizens have very similar opinions in this dimension, so the two clusters join, leading to consensus of the population.

Last, we show a simulation setting in which party competition does not affect the

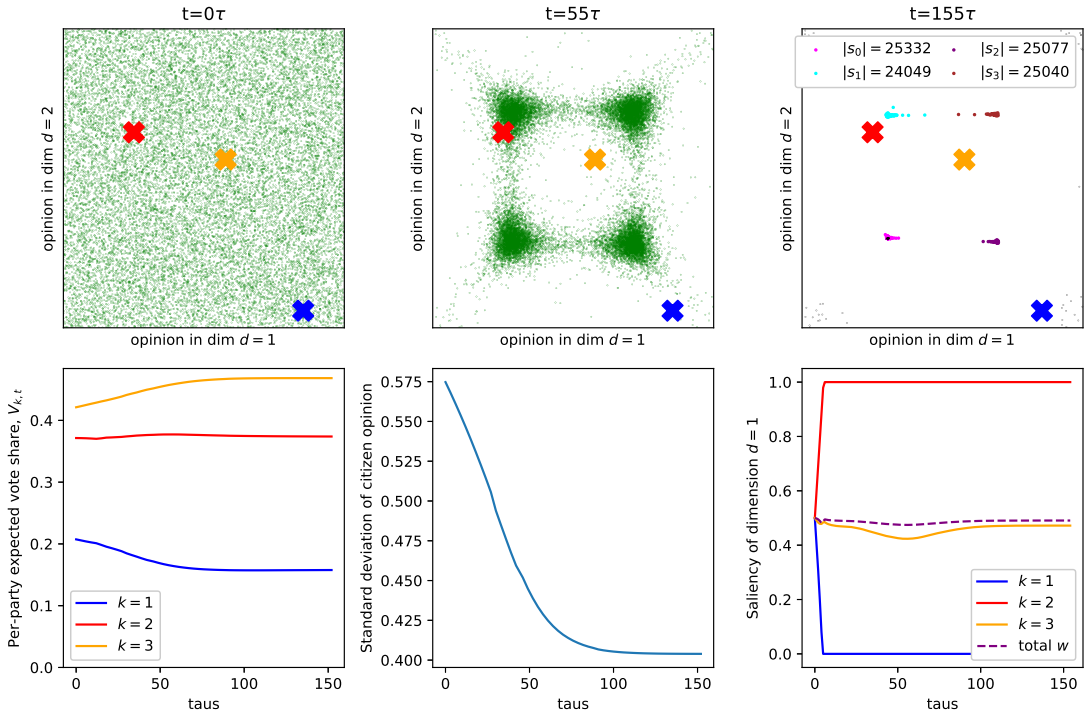


FIGURE 5.5: Simulation run for a party configuration $\boldsymbol{p}^N$ with no effect in the opinion dynamics for $\delta = 0.195$. The first row shows the initial conditions (left column), a middle snapshot of the simulation (middle column), and the stationary outcome (right column) for runs without (top row) and with (middle row) party competition. Coloured crosses indicate parties' opinions and the shown area covers the whole opinion space O. In the top–right panel, opinions are coloured by the cluster they belong to, while legends show cluster sizes. The bottom row shows the evolution of each party's expected vote share (left), dispersion in citizens' opinion (middle), and parties' promotion of dimension $d = 1$ (right).

opinion dynamics, shown in Fig. 5.5, with $\delta = \delta^{**} = 0.195$, party configuration pN, and same seed of the random number generator as in Figs. 5.2, 5.4, and 5.3. We omit again the simulation run without party competition, shown at the top row of Fig. 5.3, where the evolution of citizen opinions is identical as when including party configuration $\boldsymbol{p}^N$ (middle row of Fig. 5.5). Regarding the evolution of party saliency in this setting (bottom–right panel), we note a full promotion of each dimension by two parties from the beginning of the simulation, while the third party promotes both dimensions almost equally for the whole simulation. Therefore, the aggregated saliency results in citizens perceiving both dimensions equally, which is an identical setting to when there is no party competition.

From these single simulation runs, we have drawn a set of hypotheses linking the different effects that party competition may have on opinion dynamics to the combined behaviours of parties in saliency promotion. The promotion of polarisation is hypothesised to be caused by a strong imbalance in saliency promotion that reduces the dimensionality of the opinion space and that remains unchanged for the whole simulation time. In contrast the promotion of consensus is a result of an interplay of

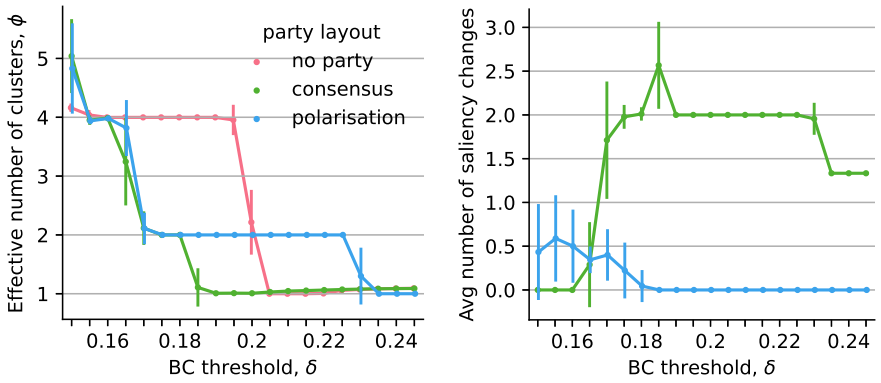


FIGURE 5.6: **a** Dependence of the effective number of clusters ϕ on the BC tolerance threshold δ for a fostering–consensus party configuration (green), a fostering–polarisation party configuration (blue), or without party competition (green). **b** Dependence of the average number of saliency changes $\langle \sigma_k \rangle$ over parties and simulation runs for the same experiments. Error bars represent the standard deviation over 30 simulation runs.

the opinion and party dynamics characterised by shifts in the saliency promotion behaviour, which in turn serves to aggregate citizen opinions in the two dimensions one at a time, i.e. is a result of the adaptive behaviour of the external controllers reacting to the dynamics of the opinions. Last, a symmetry in the saliency promotion attempts cancels out their aggregated effect, resulting in no effect in the dynamics of citizen opinions.

While the particular simulations shown above allow us to understand the different effects that party competition emerging from the proposed model, the analysis is restricted to single runs. One may wonder how robust the observed effects are with respect to the stochasticity of the model. To verify the generalisation of the hypothesis regarding the behaviours in saliency promotion, we extend the analysis to other values of δ and calculate averages over multiple runs of the opinion dynamics. Fig. 5.6**b** illustrates the dependence of the mean effective number of clusters ϕ on the tolerance threshold δ for the party configurations explored above. The figure shows that the number of changes remains consistent along the phases of Fig. 5.6a. Indeed, the polarising behaviour from Fig. 5.2 corresponds to *zero* changes in the direction of saliency updates (blue line). In contrast, undergoing two or more changes in direction of saliency update (green line) results in consensus promotion.

5.6 Party configurations that promote consensus or polarisation

Above, we have seen how the party behaviour can affect opinion dynamics and the outcome of the consensus–forming process. Here, we look one step further by

inspecting how the characteristics of the party configuration are linked to the different party behaviours in saliency promotion that in turn affect opinion dynamics. To investigate the link, we take the 500 party configurations generated in the experiments of Sect. 5.4 and label them according to whether the effect they have in the opinion dynamics: 'fostering polarisation' (351 configurations), 'fostering consensus' (86 configurations), or 'no effect' (63 configurations). We then run a decision tree algorithm to classify them based on the party system features and inspect the decision rules generated by the trees.

The resulting tree has an accuracy of 56.1%. To find where the decision tree misperformed, we can inspect the confusion matrix that compares how the classes that model predicted only from the features of the party system relate to the real classes resulting from the simulations (Fig. 5.7**a**). The confusion matrix shows that when the model uses the features to predict a fostering of polarisation, these are indeed party configurations that foster polarisation (in around 90% of the cases). Therefore, the rule dictating that the centre of mass of the triangle must have a distance to the centre of the space smaller than 0.405 (left–most box of Fig. 5.7**b**), is able alone to identify configurations that foster polarisation with a high precision. A further second rule dictating that the two smaller angles of the triangle should be smaller than 41.3º (i.e. the largest angle is larger than 97.4º), favouring obtuse–angled triangles.

On the contrary, as the decision tree tends to misclassify fostering consensus or not affecting the opinion dynamics, giving these labels more often that not to party configurations whose effects differ (left and middle columns of Fig. 5.7**a**). Therefore, the rules to identify these effects should be disregarded as not informative (upper branch of the tree in Fig. 5.7)

We can conclude that party systems (of three parties) that are well centred around the space and that form an obtuse–angled triangle tend to foster polarisation. This distribution seems to generate an imbalance of saliency promotion, with two of the three parties promoting one of the dimensions and maintaining this behaviour throughout the whole process.

5.7 Extending to other numbers of parties

Experiments so far have only covered the specific case of $K = 3$ parties, where party competition mostly has a fostering–polarisation effect, along a significant percentage of configurations fostering consensus or without having effect on the opnion dynamics. In this section, we explore party systems with a range of other numbers K of parties, from $K = 2$ to $K = 7$. For each K, we randomly sample 500 party configurations and find the location of the two phase transitions $\delta_{\phi:1\to2}$ and $\delta_{\phi:2\to2+}$ (as in Sect. 5.4) and classify their effect in opinion dynamics according to $\delta_{\phi:1\to2}$, with a

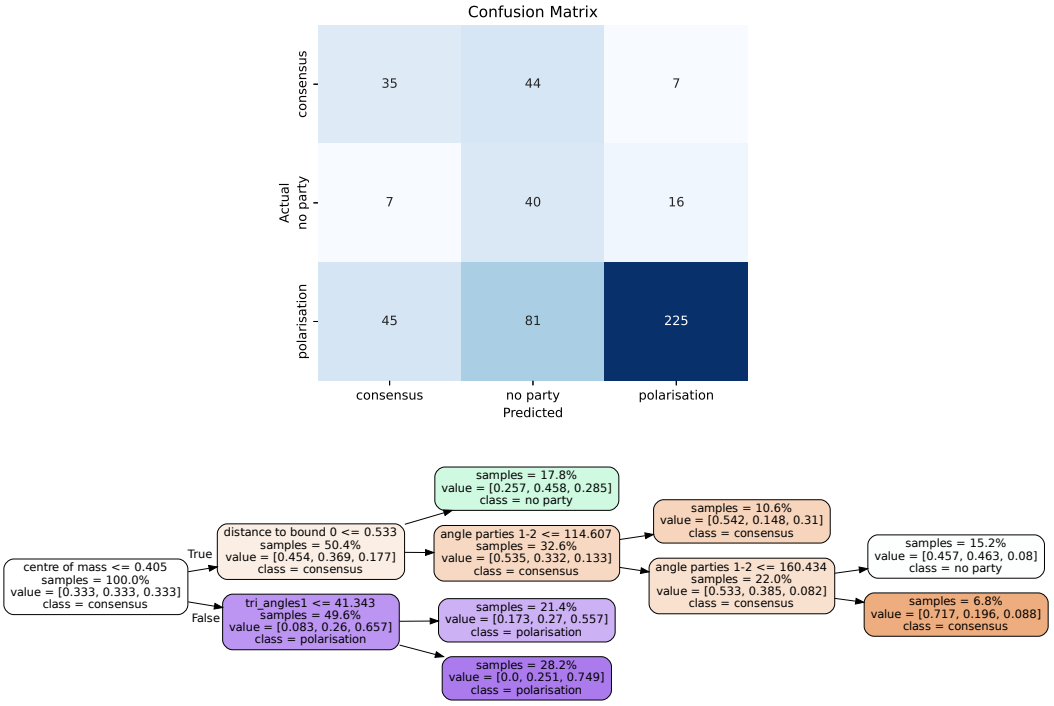


FIGURE 5.7: **a** Confusion matrix of the decision tree comparing showing the combinations of predicted effects (columns) and actual effects (row) of party configurations in the opinion dynamics. **b** Decision tree structure. The top line of each branching box shows the decision rule for branching, with the outgoing top branch corresponding to the condition satisfied and the outgoing bottom branch otherwise. Shown significant features are the distance of the centre of mass to the centre of the space (*'centre of mass'*), the amplitude of the second smaller angle (*'tri_angles1'*), the distance to the boundary of the space by the closest party to it (*'distance to bound 0'*), and the second smallest angle between two parties if in polar coordinates (*'angle between parties 1-2'*). Other lines in each box contain the percentage of party configurations within the group (*'samples'*), the weighted fraction of each class within the group (in the following order: *consensus*, *no change*, *polarisation*), and the predominant class in the group (*class*). The total tree accuracy is 56.1% out of 500 samples.

value of $\delta_{\phi:1\to2} < 0.2$ implying a fostering–consensus effect, $\delta_{\phi:1\to2} > 0.205$ a fostering–polarisation effect, and $\delta_{\phi:1\to2} = 0.2025$ no effect in the opinion dynamics. Phase diagrams for each K can be found in Appendix C.2.

Figure 5.8 shows how the presence of these three outcomes varies with the number of parties K present in the political system. We note that polarisation (orange line) is the most common outcome, happening more than 50% of the time and varying little with K, except when $K = 2$. The case of $K = 2$ parties stands out for rarely affecting the opinion dynamics (85% of the party configurations have no effect). This behaviour is related to the mechanisms shown in Fig. 5.5, presenting a very balanced saliency aggregation, which is a much easier outcome to achieve with the higher symmetrical scenario of having two parties. Indeed, as party competition is a zero–sum game and parties in our model having equal power to affect saliency it is usual that parties

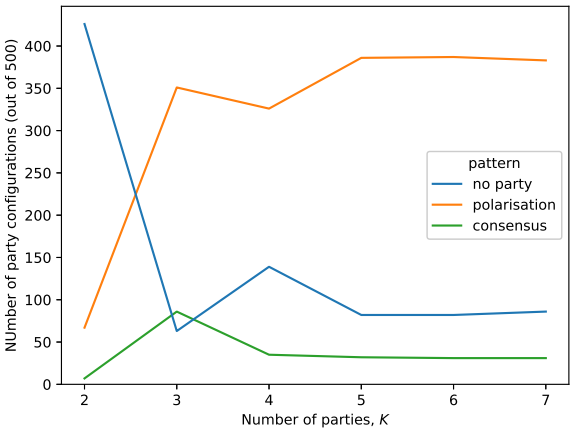


FIGURE 5.8: Proportion of party configurations producing each of the three possible effects: no effect on the opinion dynamics (blue), fostering consensus (green), and fostering polarisation (orange).

support different dimensions. This effect is repeated in a weakened form when $K = 4$ (28% of the time), as again parties may split in a symmetrical 2-vs-2 behaviour, although the presence of more parties renders this outcome more unlikely (around 16% of the time). Remarkably, outcomes of consensus promotion (in red) have a very mild presence (around 6%), only modestly more significant for $K = 3$ (17%).

5.8 Including social network structure

All experiments above assume that citizens can communicate with any other citizen indistinctly, an particularly unrealistic assumption when considering large populations, as there are clear limits in how many different others an individual can interact with (Tamarit et al., 2018). In this section, we explore a more realistic scenario in which citizens are embedded in a social network and only interact with a small set of close connections ('neighbours' in the graph terminology) and how this sparser social network structure affects the opinion dynamics and the effects that party competition has on them. As a network architecture, we use Erdos–Renyi random networks, which are commonly used in agent–based models of opinion dynamics (Amblard et al., 2015), and which can substantially increase the sparsity of the connections while keeping a similar overall structure than the complete network (Newman, 2003). We build the network as to produce an expected node degree of $\langle d \rangle = 10$, which is a realistic connection pattern in real–world scenarios.

Figure 5.9**a** shows the dependence of outcomes on the tolerance threshold δ for three cases: one without party competition, one with a party configuration that fostered consensus in the all–to–all case, and one which fostered polarisation in the all–to–all

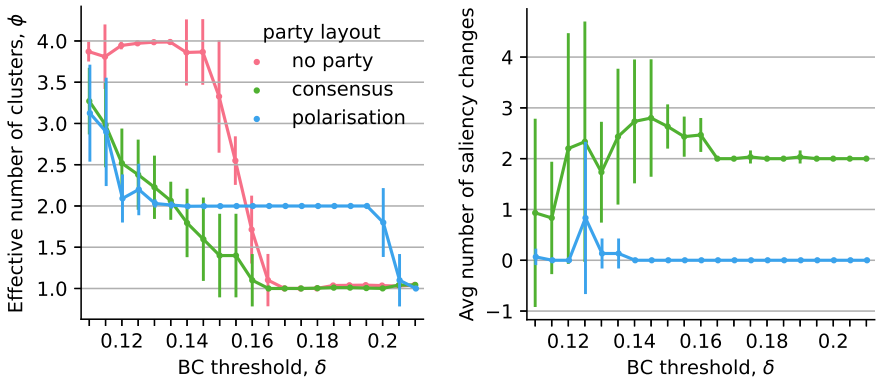


FIGURE 5.9: **a** Dependence of the effective number of clusters ϕ on the BC tolerance threshold δ for a fostering–consensus party configuration (green), a fostering–polarisation party configuration (blue), or without party competition (green). **b** Dependence of the average number of saliency changes $\langle \sigma_k \rangle$ over parties and simulation runs for the same experiments. Error bars represent the standard deviation over 10 simulation runs.

case. The first finding that stands out is a general shift of all phase transitions to lower values of tolerance threshold δ, implying that consensus can be achieved with lower values of the tolerance threshold. For instance, when there is no party competition (pink), the transition from consensus to fragmentation is at $\delta_{\phi:1\to 2} = 0.1625$, while it was at $\delta_{\phi:1\to 2} = 0.2025$ in the all–to–all scenario. Similar shifts in the tolerance threshold are apparent for the two party configurations tested, although the inclusion of social network did not alter their effect of fostering consensus and polarisation, respectively.

Figure 5.9**b** shows the average number of saliency changes per party and simulation $\langle \sigma_k \rangle$ for each of the two party configurations and for different values of δ. As in the all–to–all case, the party configuration fostering consensus showcased various changes in their saliency promotion strategies, while the party configuration fostering polarisation remained static in its saliency strategy.

5.9 Discussion

In this chapter, we have studied the interrelated dynamics of party competition via saliency adjustment and opinion dynamics in the form of a bounded–confidence model, thus exploring the feedback dynamics of adaptive, external controllers and opinion dynamics that act in similar time scales. In an exploration of different parameter regimes, we have pointed to three possible effects that party competition may have on consensus formation depending on the number of competing parties and their configuration. The first effect refers to party competition fostering a polarised state of the society. This outcome results from parties chiefly promoting the same

political dimension for the whole opinion formation process, making the electorate more divided along this dimension. In contrast, the second effect is the fostering of consensus, which happens when parties alternate saliency promotion between different political dimensions. Such behaviour initially creates division along one political dimension, and then withdraws importance from it, allowing the clusters of opinion to merge. Last, the third effect is a lack of effect, with parties saliency promotion aggregating to a balanced saliency mix that is equivalent to the unaltered state of the system.

The presence of these effects, however, is asymmetric and dependent on party configuration and numbers. While the promotion of polarisation is consistent in most studied cases, consensus fostering is a rarer effect, happening slightly more often when $K = 3$ parties are competing. We found that a party configuration of three parties whose centre of mass is near the centre of the opinion space is linked to a fostering–polarisation effect, and even more if the triangle they form is obtuse–angled. On the contrary, the presence of only $K = 2$ parties competing is particular in that it has negligible effects on the opinion dynamics due to the symmetry of the scenario that crates a zero–sum game that causes their actions to cancel out. This symmetry effect is also seen at a much smaller scale when $K = 4$. An interesting modification that would break this asymmetry and potentially arrive at other interesting phenomena is changing the saliency aggregation function to possibly make a party's contribution to the total saliency proportional to their attained vote share, or to their distance to the citizen's opinion.

Embedding the population into a sparse social network has the effect of shifting the phase transitions to lower values of the tolerance threshold implying that consensus is more easily achieved with sparser social connections. This effect has also been previously observed in 1D bounded–confidence models and it has been linked to the presence of bridges that facilitate the interaction of clusters, in both uniform random graphs and scale–free networks (Schawe et al., 2021). However, the introduction of a social structure does not seem to alter the effects that party competition have in the opinion dynamics once the shift to lower values of the threshold is accounted for.

There are many possible extensions to this work, as there are many ingredients in the model that could be modified or further explored. We expose here a few of particular interest. First, the examination of which features of the party system configuration are responsible for each effect can be extended. Although decision trees are exceptional in their interpretability, their approach is a greedy one — selecting one feature at a time to compute each split — so their solutions are not necessarily optimal and are unable to express concepts such as XOR (Krzywinski and Altman, 2017). However, maybe the effects are better described by combinations of several features, which would require more sophisticated approaches. Second, we have studied a scenario with only two dimensions. The case of a three–dimensional opinion spaces is also worth exploring,

as political spaces with three dimensions is also common among Western democracies (Reiljan et al., 2020). Also, we have assumed static scenario, where the number of parties and dimensions are fixed from the beginning of the simulation. However, it would be interesting to explore the appearance of a new dimension or a new party some time after the simulation has started, which is a scenario common in the emergence of populist parties (Tavits, 2008). Last, the inclusion of turnout and abstention in voting — when a citizen has no party near her in the opinion space — would be of particular interest (Fowler and Smirnov, 2005), as well as allowing parties to move their positions (with restrictions) in the opinion space (Fowler and Laver, 2008).

Furthermore, we have made a number of choices in the model with potentially strong impacts on the results. For instance, we have assumed a 2-norm Euclidean metric as the distance metric, while some other works in bounded–confidence models have employed the 1- or the ∞- norms instead (Lorenz, 2008; Fortunato et al., 2005). Likewise, we have assumed that all parties have equal power in affecting the perceived saliency of the population and that their efforts are aggregated linearly, which are both unrealistic assumptions. It would be particularly interesting to couple the influence power of a party to its instantaneous vote share or to the distance in opinion space to the citizens they are affecting. Additionally, other mechanisms of update of party saliency can be explored, such as stochastic hill–climbing (Kollman et al., 1992).

Chapter 6

Summary and conclusions

This thesis has advanced the understanding of how external control interacts with opinion dynamics under scenarios with multiple external controllers, developing and analysing strategies of optimal targeting, and studying their interference with opinion dynamics processes when they react adaptively. We have first studied the scenario of external controllers as 'perfect optimisers' that attempt to promote one of two choices in a social group embedded in a network of interactions. Under the assumptions that controllers have full knowledge of the context, that they compete against an opponent, and that they can spread their influence attempts continuously among the individuals in the social group, we have developed a gradient–ascent algorithm that provides optimal solutions to the influence maximisation problem of optimally distributing the influence, with guarantees that a very good approximation of the optimal solution is found in polynomial time. This is a major advantage over influence maximisation when formulated with discrete influence targets — which is an NP-hard problem (Kempe et al., 2003), and much research has focused on finding approximations that obtain good enough solutions in reasonable time (Li et al., 2018).

More importantly, we have provided a detailed intuition on what aspects of the influence strategies make them successful when facing an opponent, in scenarios with different levels of power imbalance. We have used an analytical and numerical exploration of the structure of optimal solutions and arrived at two heuristics that can explain a big part of the observed optimal allocations. These heuristics — that we term *shadowing* and *shielding* — correspond respectively to first– and further–order reactions to the opponent's behaviour — where order refers to the distance to the nodes targeted by the opponent. Both heuristics depend greatly on the strategy of the opponent and the ratio of influence budgets, and only marginally on structural properties of the nodes, such as node degree. Notably, the shielding heuristic explains well why nodes with a high degree tend to be good targets for influence maximisation, as these nodes have higher chances of being neighbours of those targeted by the opponent. Therefore, a basic degree–based heuristic (Dezső and Barabási, 2002) would

be sensible if the active controller does not have any information about the allocations of the opponent, which goes in line with much of the previous literature. However, we have also shown that shadowing and shielding only partly explain optimal allocations in the classical approach with discrete allocations, where the degree–based heuristic has a proper contribution on its own (i.e. beyond shielding) to optimal solutions. This detailed analysis of the structure of optimal solutions is absent in most previous work, which mainly focuses on developing algorithmic solutions to the problem (Li et al., 2018; Kempe et al., 2003; Yadav et al., 2017; Lynn and Lee, 2018).

We have then studied an extension of the above scenario where the social group is heterogeneous and individuals may be biased, i.e. they prefer one of the opinions, which is adopted with higher probability (Masuda et al., 2010). We have extended the question of optimal targetting to this context and studied when it is strategically optimal to target the biased agents or the unbiased ones. By adapting the gradient–ascent algorithm to this scenario, we have been able to analyse its optimal solutions and showed that targeting individuals biased against the external controller is only optimal if the controller has an advantageous position of high influence budget (and higher than the opponent) or an aggregate bias of the social group favouring her target opinion; while if in a disadvantageous position optimal strategies target individuals "on her side", i.e. biased against the opposing opinion. Furthermore, optimal controllers with large budgets allocate more influence to high–degree nodes (hubs) while preferring low–degree nodes (periphery) when the available budget is small, creating an interplay between node degree and the level of bias.

Last, to cover different context and assumptions of external control, we have shifted the focus to controllers as 'imperfect optimisers', i.e. having limited information about the social dynamics, its structure, or the strategy of the opponents, so that they can only react optimally to the instant state of the system in a form of adaptive control. More specifically, we have drawn inspiration from the context of political parties that campaign for votes while the population of citizens follows a consensus–formation process, where opinions are placed in a multidimensional political space, and with both dynamics occurring at similar time scales. We have studied the effects that these adaptive controllers may induce on the dynamics of the population when myopically seeking votes, and found that their interference can foster polarisation of opinions, although, surprisingly, they can also have the opposite effect and assist in the arrival at a consensus. The effect produced depends on the configuration of the party system, such as the number of parties and their spatial distribution, with the promotion of polarisation being a much more prevalent and consistent outcome than of consensus. When the party system involves three parties, the promotion of consensus is more likely when they are well centred in the political space and form an obtuse–angled triangle. If the population is embedded in a sparse, random social network, it is more prone to reach a consensus state, but the effects that party competition has on this

process remain qualitatively similar than in the scenario where any citizen can interact with any other. The role of external controllers affecting the importance of different opinion dimensions when perceiving distances has not previously been studied, which has a relevant role in the political realm in the context of voting (Dragu and Fan, 2016; Feld et al., 2014; De Sio and Weber, 2014; Amorós and Puy, 2010). We provide the first account of this setting and its relation to consensus processes.

In summary, we have studied external control in a series of scenarios and assumptions to better understand its optimality and effects on opinion dynamics processes and in the presence of multiple controllers. Our findings provide tools to obtain optimal targetting strategies, as well as intuition on how these look like in the form of heuristics. These heuristics not only serve as shortcuts to arrive at good enough solutions, but they can also be templates that facilitates comparison of optimal strategies to other scenarios and model assumptions. The tools and findings provided can be applied to diverse contexts, such as policy making, political scenarios, marketing, or technological promotion and epidemic control, where they can be useful to bring particular states to social groups, combat an opponent doing so, gain insight on how to prevent external influence to be effective, or be aware of the side effects that external control may have in the opinion dynamics.

6.1 Critical evaluation and future work

Although significant, our contribution is a small drop in the face of the challenges that the study of external control brings. This is because it is an incredibly complex phenomenon for which there are clear limits in the access to information on its working — opinions are not easy to conceptualise and measure, and particularly more during the longer period of times required to probe their evolution and at the same time as the trace of interactions — as well as from the external controllers — they tend to be rather secretive and avoid reporting their strategies and their findings. Furthermore, it the phenomenon studied calls for a highly multi-disciplinary approach, as it requires understanding of phenomena at different levels: cognitive processes that trigger when adopting new opinions (cognitive science), patterns of interactions in populations (social sciences), how external controllers operate (marketing and political science), how micro-behaviours can result in emerging phenomena (complex systems), optimisation tools to find best strategies (computer science), and how competing agents arrive at a joint optimal strategy (game theory). This thesis is mostly based on research and tools from computer science and complex systems, while drawing insights from the previous ones. However, fully understanding the problems from all the perspectives is outside the scope of this thesis, and can be better approach with a multi-disciplinary team, which may be able to go beyond some of the limitations of this work, which are listed in the following.

First, we have attempted to cover different assumptions and scenarios, while at the same time we attempted to keep them as abstract as possible with the aim to preserve the generality of the methods and findings and their applicability. However, this generality risks blurring the link to real applications. Therefore, it would be important to probe the validity of the findings by linking them to specific instances of external control cases for which data is available. This can include e.g. cases of elections, for which there is data about the campaigns in the media and polls regarding the voters' preferences (Moya et al., 2017; Sobkowicz, 2016).

Related to the previous concern, the choices taken to model the opinion dynamics processes may have fundamental effects in the behaviours emerging (Sobkowicz, 2015; Flache et al., 2017). Models are only useful if they are simple yet credible, and convey the fundamental mechanisms responsible for the research phenomenon studied. Striking this balance requires expertise and insights from application domain (Mueller and Tan, 2018; Schweighofer et al., 2020b). In this work, we have spared ourselves such a challenge by relying on widely accepted and studied models from the opinion dynamics field. We have chosen two paradigmatic models from the discrete and continuous opinion families — the voter model and the bounded–confidence model, respectively. However, other suitable models could also include further mechanisms, and it may be important to account for them. One of such mechanisms is *complex contagion* (Alshamsi et al., 2018b). by which the probability of adopting an opinion behaves non-linearly on the number of neighbours holding it, a property that is held by models such as the *majority rule* (Galam and Jacobs, 2007; Nguyen et al., 2020), the *q-voter model* (Castellano et al., 2009b), or the Glauber dynamics of the Ising model (Ising, 1925). It would be interesting to investigate whether our findings hold under this family of opinion dynamics models.

Furthermore, we have employed a variety of complex network topologies that are commonly used in the study of social phenomena and that reproduce some of the known characteristics of real social networks (Newman, 2003; Amblard et al., 2015). However, richer representations of the social interactions could be employed. For instance, the experiments and findings could be expanded to directed networks with asymmetric effects in opinion interactions, multi-layer networks that convey parallel interactions of different nature, higher–order networks that capture interactions involving multiple individuals, temporal networks where interactions are time stamped, or dynamical network with an evolving architecture. Although we do not expect our findings to change much under these scenarios, they should be tested and differences in findings accounted for.

Further limitations of the current work come from other assumptions in the studied scenarios. For instance, in the first scenario we have assumed that optimisers are 'perfect', i.e. they have full information of the social structure and the opinion dynamics. Although this is useful in that it has been used to derive heuristics that do

not need full information to generate optimal strategies, scenarios with partial information would deserve further study (Cai et al., 2022). A second assumption relates to the aim of the perfect optimisers to maximise their influence in the long–run (or any time scale larger than that of the opinion dynamics) while in many situations they are constrained by a very limited time frame in which the need to achieve their goals (Brede et al., 2019a). Such scenario introduces additional complexities in the computation of an optimal strategy (Cai et al., 2021), although in the limit of very short time horizons the scenario may be similar to that of adaptive controllers studied in the thesis.

Regarding the research problem, other interesting questions could be approached using the models and tools developed here. One important question is how can social dynamics be modified to make the social group more resilient to external control. This can be approach by examining which topological features of a social network — such as degree heterogeneity, degree assortativity, or community structure — facilitates or hinder the controllability of the social group, or how modification in the patterns of interactions or in the rules of the dynamics can achieve similar goals (Lorenz and Urbig, 2007; Li et al., 2020). A related problem with similar modelling requirements is that of *sensing enhancement*, by which the social group aims to follow an external signal (the 'truth') and alters its patterns of interaction with respect to the external signal to collectively become robust to errors in its perception (Brede and Romero-Moreno, 2022)

In conclusion, opinion dynamics and its derivations is still an area of research that offers much space for exploration, while highly relevant in the wake of recent social phenomena. Indeed, the research field is currently highly active and progressively enriching from the collaboration of various disciplines, bridging the gaps from the social and psychological sciences (Sobkowicz, 2015; Mueller and Tan, 2018; Schweighofer et al., 2020b). Linking the profusion of studies in models of opinion dynamics to specific and real scenarios would further benefit the understanding of the behaviour of external control in social groups. Furthermore, as data about social interactions is produced at higher rates and quantities, its availability may facilitate the task of model validation and the use of data–driven approaches. Therefore, research in the field can be expected to grow in quantity and quality in the near future.

Appendix A

Appendix to Chapter 3

A.1 Testing the HMF approximation

In the following, we consider random networks with bimodal degree distribution as a simple class of leader–follower type networks. Leader–follower topologies are frequently found on different social formations, such as social media (Rathnayake and Suthers, 2016) or corporate hierarchies (Tichy et al., 1979). We employ synthetic random networks that contain nodes with only two possible degrees: d_1 (*followers*) and $d_2 > d_1$ (*leaders*), whose proportions are denoted by ρ and $1-\rho$, respectively. Varying the values of d_1 and d_2 allows us to control the connectivity and heterogeneity of the resulting networks. We also introduce influence allocation variables, $\alpha \in [0,1]$ for the A- and B-controller, respectively, that reflect the percentage of the budget allocated to low-degree nodes, with $1-\alpha$ and $1-\beta$ corresponding to the percentage of the budget that each controller allocates to nodes with high degrees. We further assume that controllers target nodes of the same group with equal strength, with allocations of the A-controller to nodes of each group corresponding to $w_{a1} = \alpha \mathcal{B}_a/\rho$ and $w_{a2} = (1-\alpha)\mathcal{B}_a/(1-\rho)$, and analogous formulations for the B-controller with β. From (3.8), the HMF approximation for the vote share in this class of networks can be computed as

$$\begin{aligned} X^{\mathrm{HMF}} = &\left(\tfrac{\rho\, d_1}{d_1+w_{a1}+w_{b1}} + \tfrac{(1-\rho)\, d_2}{d_2+w_{a2}+w_{b2}}\right) \times \left(\tfrac{\rho\, w_{a1} d_1}{d_1+w_{a1}+w_{b1}} + \tfrac{(1-\rho) w_{a2} d_2}{d_2+w_{a2}+w_{b2}}\right) \times \\ &\times \left(\tfrac{\rho(w_{a1}+w_{b1}) d_1}{d_1+w_{a1}+w_{b1}} + \tfrac{(1-\rho)(w_{a2}+w_{b2}) d_2}{d_2+w_{a2}+w_{b2}}\right)^{-1} + \left(\tfrac{\rho w_{a1}}{d_1+w_{a1}+w_{b1}} + \tfrac{(1-\rho) w_{a2}}{d_2+w_{a2}+w_{b2}}\right) . \end{aligned} \tag{A.1}$$

In this reduced setting, the controller has a single degree of freedom, α, and needs to solve $\partial X^{\mathrm{HMF}}/\partial\alpha = 0$. The solution to this high–order polynomial, α^*, can be found via numerical methods.

To validate the accuracy of the approximation, we compare the vote shares obtained via (A.1) to exact results calculated directly from (3.2) on leader–follower networks.

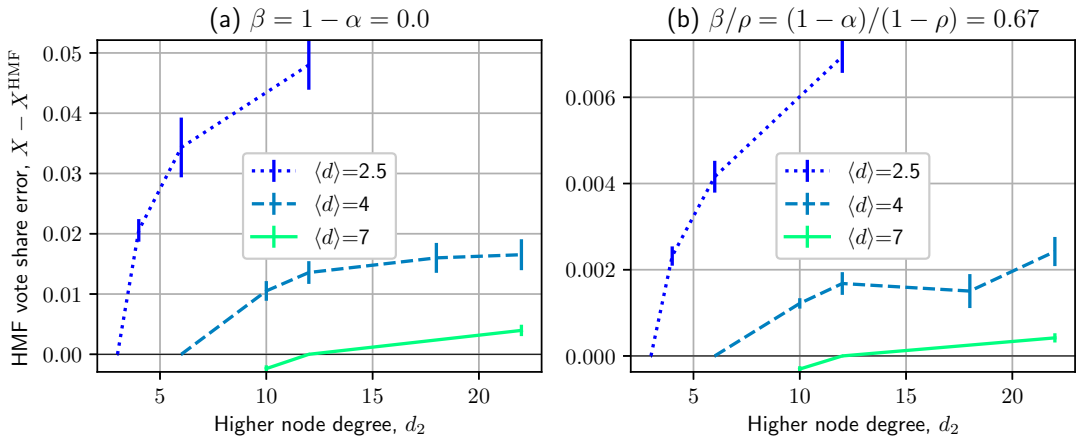


FIGURE A.1: Errors in vote share of the HMF approximation with respect to the exact solution on leader–follower networks with fixed low degree $d_1 = 2$, and varying mean ($\langle d \rangle$) and high (d_2) degrees. Budgets are the same for both controllers and equal to $N\langle d \rangle/3$. Error bars correspond to the standard deviation of errors across 15 randomly generated networks of size $N = 1000$. In panel (a), the A-controller only targets high–degree nodes, while the B-controller only targets low–degree nodes. In panel (b), link weights from the A-controller to high–degree nodes are 33% less strong than weights given if allocations were uniformly distributed. Similarly, low degree nodes receive from the B-controller 67% of the strength they would receive if allocations were uniform across all nodes.

Random networks of size $N = 1000$ are sampled from configuration models (Bayati et al., 2010) that enforce ρN nodes to have degree d_1 and $(1-\rho)N$ nodes to have degree d_2. To preserve the degree of connectivity across different networks while varying the degree heterogeneity, we fix an average degree $\langle d \rangle = \rho d_1 + (1-\rho)d_2$ and make joint changes in ρ and d_2. We consider controllers that target all nodes from each group with equal allocation. Due to general symmetries, only two dynamic degrees of freedom remain in the system, $x_{d=d_1}$ and $x_{d=d_2}$ (x_1 and x_2 from now on).

To reduce the exploration space, we assume that both controllers have equal budgets $\mathcal{B}_a = \mathcal{B}_b$ and —to create some asymmetry— we assume they target high– and low–degree nodes differently. Specifically, the A-controller targets low–degree nodes with higher strength than high–degree nodes ($\alpha > \rho$), and the B-controller targets high–degree nodes more strongly than low–degree nodes ($\beta < \rho$).

Figure A.1 compares mean–field vote shares X^{HMF} obtained via (A.1) to the exact solution obtained from solving (3.2). Experiments are performed with low–degree $d_1 = 2$, budgets $\mathcal{B} = N\langle d \rangle/3$, and different combinations of high degree d_2, average degree $\langle d \rangle$, and budget allocations α, β. As it can be seen in Fig. A.1a, errors in the HMF approximation increase as $\langle d \rangle$ decreases —i.e. the sparser networks are. While the highest error for $\langle d \rangle = 2.5$ reaches a magnitude of 0.05 in vote share, errors drop below 0.02 in vote share for $\langle d \rangle = 4$ and are under 10^{-3} for mean degrees larger than $\langle d \rangle = 7$. When comparing experiments with the same $\langle d \rangle$ (across the same line), we

observe that the approximation has higher accuracy for networks with lower levels of heterogeneity (low d_2), with errors becoming negligible as d_2 approaches $\langle d \rangle$.

While in Fig. A.1a the A-controller focuses all its weight on low–degree nodes and the B-controller exclusively focuses on high–degree nodes, in Fig. A.1b we explore a more balanced budget allocation, with $\beta/\rho = (1-\alpha)/(1-\rho) = 0.67$. This more nuanced setting qualitatively preserves the effects that $\langle d \rangle$ and d_2 have on errors, but improves the approximation accuracy in all scenarios, drawing down the maximum error of $\langle d \rangle = 2.5$ to around 0.007 in vote share, and the error of higher values of $\langle d \rangle$ to around 0.002 vote share and below.

Results from these experiments demonstrate that errors of the HMF approximation on leader–follower networks stay below 0.02 vote share for networks with $\langle d \rangle \geq 4$, while they can be $O(10^{-3})$ when controllers do not concentrate all their budget on a subset of nodes in the network. Although not shown in the figures, changing other parameters (e.g. increasing d_1, or $\mathcal{B}_a$ and $\mathcal{B}_b$) can bring the upper bound of errors further down by one order of magnitude.

A.2 Testing the GA algorithm for IM

In this section, we first present a single run of the GA algorithm on a leader–follower network and compare its optimal allocations to analytical solutions obtained from the HMF approximation. We will be using X^{GA} and X^{ana} to refer to exact vote shares (obtained via (3.2)) when the A-controller uses optimal allocations obtained analytically (X^{ana}) or via gradient–ascent (X^{GA}).

Figure A.2 shows different aspects of a sample run of the gradient ascent algorithm. Figure A.2a gives the vote share increase through the iterations of the run. At the first iteration, the A-controller starts distributing the budget equally among all nodes. The difference in resulting vote share between this initial strategy, with $X_0^{\mathrm{GA}} = 0.2355$, and the final strategy found by the GA, with $X^{\mathrm{GA}} = 0.2502$, accounts for $X^{\mathrm{GA}} - X_0^{\mathrm{GA}} = 0.0148$. We also find that the exact vote share of the optimal allocation given by the analytical solution, $X^{\mathrm{ana}} = 0.2501$, is almost identical to the GA solution, $X^{\mathrm{GA}} = 0.2502$, and one order of magnitude lower than errors in vote share of the HMF approximation: $|X^{\mathrm{ana}} - X^{\mathrm{ana}}_{\mathrm{HMF}}| = 0.0022$. Fig. A.2b shows the optimal allocation found by the algorithm for every single node. Nodes with identical degrees receive similar allocations, which is in good agreement with the assumption of uniform spreading across nodes of the same group, although small fluctuations resulting from different topological positions of the nodes in the network can be noted. Fig. A.2c illustrates the working of the gradient ascent algorithm. The axes of this sub-figure represent the average influence allocation to low (d_1) and high (d_2) degree nodes. The green line gives the N-simplex constraint. Blue points represent the steps taken by the GA

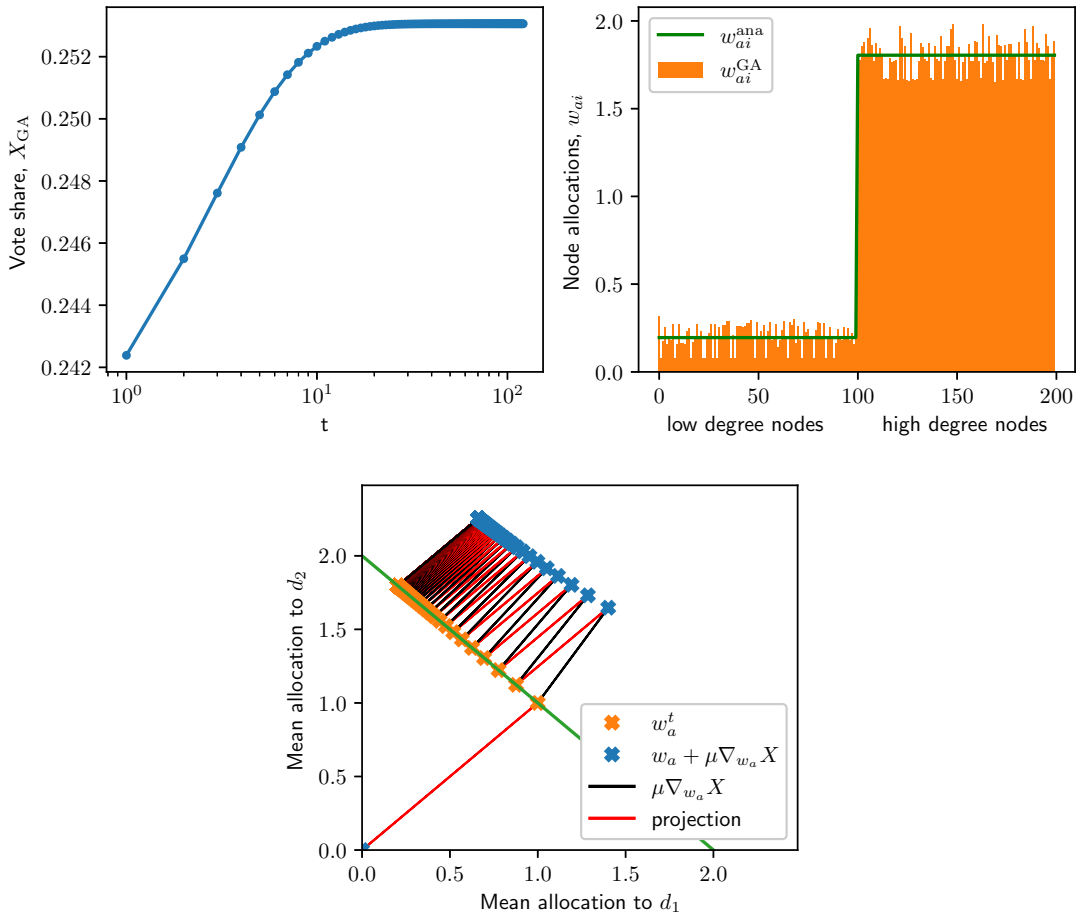


FIGURE A.2: (a) Improvements in vote share throughout the iterations (on log scale), (b) resulting individual node allocations and (c) gradient steps of a single sample run of the gradient ascent algorithm. This sample run is performed on a network of size $N = 200$, fraction of low–degree nodes $\rho = 0.5$, low degree $d_1 = 2$, high degree $d_2 = 6$, budgets $\mathcal{B}_a = 1$ and $\mathcal{B}_b = 4$, allocation to low–degree nodes by the B-controller $\beta = 0.8$, and step size of the GA $\mu = 3$. The solid green line in (b) indicates the optimal allocation from the analytical solution.

algorithm and orange points give the projections back onto the simplex, with projecting lines in red. The final fraction of the budget allocated to low–degree nodes by the GA algorithm, $\alpha_{GA} = 0.0988$, barely differs from that given to low degree nodes by the analytical solution, $\alpha_{ana} = 0.0977$.

Figure A.3 compares analytical and GA optimal allocations for networks with bimodal degree distribution with varying network parameters $\langle d \rangle$ and d_2. Based on the approximation errors shown in Sec. A.1, the network characteristics have been chosen for low errors of the approximation, thus ensuring accurate analytical solutions to optimal allocations. Figure A.3a illustrates the gain in vote share when using the GA solutions as compared to analytical solutions, i.e. $X^{GA} - X^{ana}$. Differences are of the order of 10^{-4} or less (for less sparse networks). Panel (b) of the same figure shows

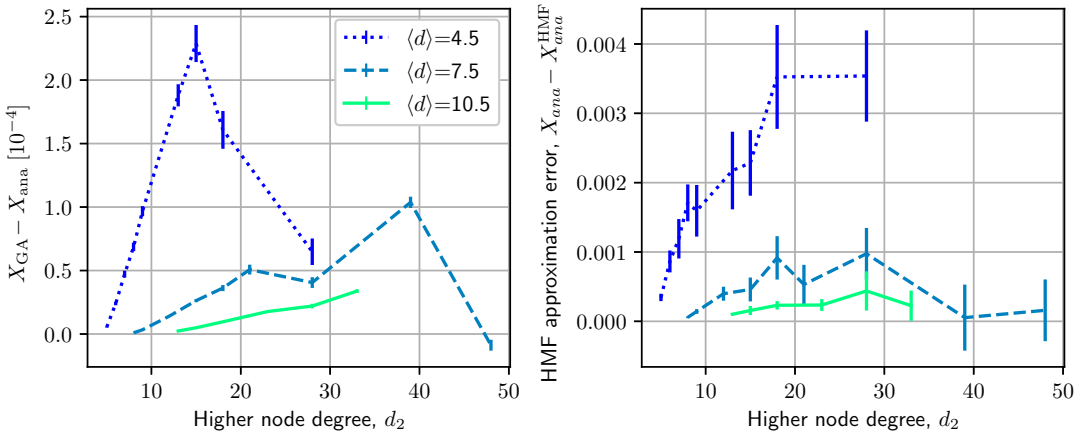


FIGURE A.3: (a) Differences in resulting vote share between analytical and numerical solutions, and (b) errors of the HMF estimation of the vote share in the analytical solution, and their dependence on degree heterogeneity (measured by d_2) and connectivity (measured by $\langle d \rangle$). Points represent averages over 15 randomly generated networks of size $N = 1000$ and error bars give standard deviations. Other parameters correspond to $d_1 = 3$, $\mathcal{B}_a = \mathcal{B}_b = 2N\langle d \rangle/3$, and $\beta/\rho = 0.5$.

errors in vote share of the HMF estimation for the analytical solutions employed in panel (a). Errors are of the order of 10^{-3} and behave as expected from the results in Sec. A.1. Solutions of optimal influence allocations determined by the two methods are in very good agreement in most scenarios, with differences of the order of 10^{-4} even when errors of the HMF approximation are of the order of 10^{-3}. However, the GA algorithm achieves a slightly better performance than the approximation in all cases. This can be explained by the capability of the GA algorithm to take into account each node's specific topological position, while the HMF approximation only works with average properties of nodes of the same degree.

A.3 Details of the network employed

For all experiments in the main manuscript, we have used an email interaction network (Guimerà et al., 2003) as an example of a common topology found in social networks. This heterogeneous network is unweighted and undirected, with a unique component of size $N = 1133$, mean degree $\langle d \rangle = 9.62$, and degree assortativity $\delta = 0.078$. Its degree distribution can be seen in Fig A.4. Experiments on other network topologies are shown below.

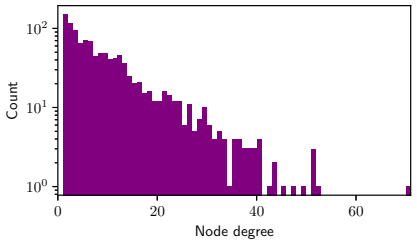


FIGURE A.4: Degree distribution of the studied network.

A.4 Numerical experiments of shielding in various network topologies

The idea behind shielding is that nodes that are direct neighbours of the nodes targeted by the opponent should receive higher allocations than other nodes in the network. The presence of shielding is then indicated by a strong dependence of allocations to nodes on the distance to nodes targeted by the opponent. We analyse here optimal allocations obtained numerically via gradient ascent on a variety of complex networks and check the presence of this effect.

For these experiments, we employ networks of size $N = 361$ with the following topologies: 2D lattices, small–world networks (Watts and Strogatz, 1998) (from a ring with connectivity $d = 4$ and a rewiring probability $p = 0.2$), Barabasi–Albert (BA) networks Barabási and Albert (1999) (with every new node wiring to $m = 1$ node), and random regular graphs with degree $d = 4$. As an initial exploration, we set the passive controller to spend the whole budget on a single node. This node corresponds to the central position for the 2D grids, the highest degree node for BA networks, and is randomly chosen for the other two network topologies.

Fig. A.5 shows average optimal allocations as a function of distance from the B-targeted node, with the top row showing scenarios where the A-controller is in budget disadvantage, while at budget advantage at the bottom row. A clear pattern stands out in the figure. We observe a clear preference for control allocations to nodes directly surrounding the B-targeted node (distance one), with allocation strengths diminishing with the distance. This shielding strategy is found independent of resource advantage or disadvantage and in all studied network topologies. Moreover, we note that non-negligible control allocations tend to be also given to nodes very far away from the B-targeted nodes —particularly visible in the case of 2D lattices— reinforcing the idea that a spread of allocations through the whole network is always beneficial.

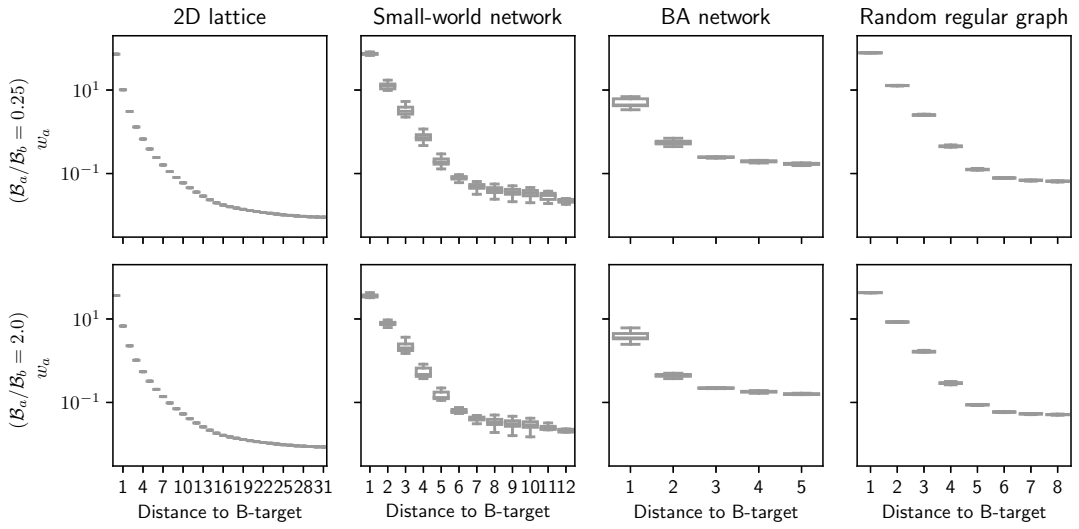


FIGURE A.5: Dependence of optimal allocations w_{ai} on the distance to the node targeted by the opponent when the opponent's budget is four times higher (top row) or half (bottom row) than that of the active controller, whose budget is fixed at $\mathcal{B}_a = 2N/3$. Boxplots gather allocations for each distance and across 15 repetitions of the experiments. Networks are of size $N = 1089$, and from left to right: 2D regular lattices, small–world networks ($k = 4$, $p = 0.2$), Barabasi-Albert (BA) networks ($m = 1$) and random 4-regular graphs.

A.5 Analytical inspection of shielding

This section provides an analytical understanding of the shielding effects observed in the numerical experiments of Sect. 3.6. For a basic model that can capture shielding while at the same time being mathematically tractable, we neglect the effects of degree heterogeneity and focus on regular random graphs; i.e. random graphs where all nodes have the same degree d. To proceed, we consider a B-controller who targets only a randomly selected fraction $0 < \rho_B < 1$ of the nodes in the network and with uniform strength $w_{bi} = \mathcal{B}_b/N\rho_B$. Below, we develop a mean–field approximation for this scenario, evaluate its validity, and compare analytical solutions to IM obtained from the approximation with numerical results obtained via gradient ascent.

A.5.1 Neighbour Mean–Field Approximation for K-Regular Graphs

Here, we recover the three disjoint groups related to shielding, $\mathbf{S} = \{\mathbf{T}_b, \mathbf{N}_b, \mathbf{R}\}$, that were introduced in Sec. 3.3. Our mean–field approximation assumes that, irrespective of exact topological positions, all nodes within the same group in $\mathbf{S}$ have identical states, and nodes of different groups are connected at random. Note that the grouping above includes nodes of distance two or higher into the same group $\mathbf{R}$, so

improvements of the method might be possible by distinguishing between second– and higher–order neighbours.

The probabilities of adopting opinion A for the nodes of each group can be defined by x_B, x_N, and x_R, respectively, with their density in the network defined by ρ_B, ρ_N, and ρ_R, subject to $\rho_B + \rho_N + \rho_R = 1$. Whereas ρ_B is a given parameter, ρ_N and ρ_R must be derived from the network topology. For our assumption of random mixing of nodes and random targeting by the controller, this calculation is simple. Nodes belonging to **R** satisfy that (1) they are not targeted by the B-controller, which happens with probability $1 - \rho_B$, and (2) none of their d neighbours is targeted by the B-controller either, which happens with probability $(1 - \rho_B)^d$. Consequently, $\rho_R = (1 - \rho_B)^{d+1}$ and ρ_N can be derived from the sum of probabilities to one: $\rho_N = 1 - \rho_B - \rho_R$.

To solve the influence maximization problem, the A-controller must decide how to optimally split her budget among the groups via determining w_{aB}, w_{aN}, and w_{aR} (i.e. the allocation strength directed to each node in $\mathbf{T}_b$, $\mathbf{N}_b$, and **R**, respectively) subject to the budget constraint $w_{aB}\,\rho_B + w_{aN}\,\rho_N + w_{aR}\,\rho_R \leq \mathcal{B}_a/N$. Using Eq. (3.2), we obtain the mean–field vote shares for the groups as

$$\begin{aligned} x_B &= \frac{d\left(\gamma_{B|B}\,x_B + \gamma_{N|B}\,x_N\right) + w_{aB}}{d + w_{aB} + \mathcal{B}_b/\rho_B}\,, \\ x_N &= \frac{d\left(\gamma_{B|N}\,x_B + \gamma_{N|N}\,x_N + \gamma_{R|N}\,x_R\right) + w_{aN}}{d + w_{aN}}\,, \\ x_R &= \frac{d\left(\gamma_{N|R}\,x_N + \gamma_{R|R}\,x_R\right) + w_{aR}}{d + w_{aR}}\,, \end{aligned} \tag{A.2}$$

where $\gamma_{Y|X}$ represents the probability for an edge to be attached to a node from group Y while the other edge is attached to a node from group X. The calculation of this probability is straight–forward for $\gamma_{Y|B}$:

$$\gamma_{B|B} = \gamma_B = \rho_B\,, \qquad \gamma_{N|B} = 1 - \rho_B\,. \tag{A.3}$$

At the other end, for $\gamma_{Y|R}$, the node at the other side of the edge will also belong to **R** if none of its other $d - 1$ edges is linked to a node in $\mathbf{T}_b$:

$$\gamma_{R|R} = (1 - \rho_B)^{d-1}\,, \qquad \gamma_{N|R} = 1 - (1 - \rho_B)^{d-1}\,. \tag{A.4}$$

Finally, for $\gamma_{Y|N}$, the use of Bayes' rule leads to

$$\begin{aligned} \gamma_{B|N} &= \gamma_{N|B}\,\gamma_B/\gamma_N &&= (1 - \rho_B)\,\rho_B/\rho_N\,, \\ \gamma_{R|N} &= \gamma_{N|R}\,\gamma_R/\gamma_N &&= \left[1 - (1 - \rho_B)^{d-1}\right]\rho_R/\rho_N\,, \\ \gamma_{N|N} &= 1 - \gamma_{B|N} - \gamma_{R|N}\,. \end{aligned} \tag{A.5}$$

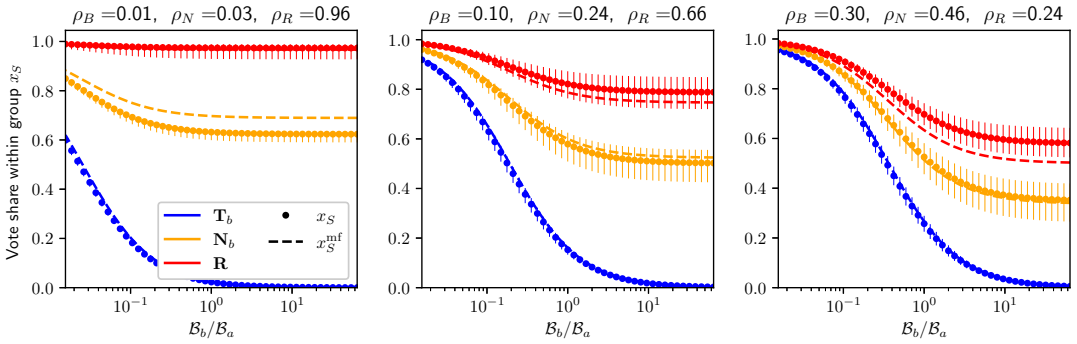


FIGURE A.6: Evaluation of the shielding mean–field approximation on 3-regular networks. Exact solutions and mean–field approximation of vote shares within the three groups in $\mathbf{S}$ for an opponent that targets a varying fraction of nodes, ρ_B, and holding a varying relative budget, $\mathcal{B}_b/\mathcal{B}_a$. Points from the exact solution, $x_{\mathbf{S}}$, are averages over 15 randomly generated networks of size $N = 1000$, with error bars corresponding to the upper and lower mean absolute deviations of vote shares within the group. Results from the mean–field approximation, $x_{\mathbf{S}}^{\mathrm{mf}}$, are given by *dashed lines*. The titles show the derived fractions of neighbours of each group present in the network. The budget for the active controller is $\mathcal{B}_a = N/2$.

Equation (A.2) can be made explicit via

$$\boldsymbol{x} = \left[\mathrm{diag}(d + \boldsymbol{w}_a + \boldsymbol{w}_b) - \Gamma d\,\right]^{-1} \boldsymbol{w}_a\ , \tag{A.6}$$

where bold symbols are vectors $\boldsymbol{y} = [y_B\ y_N\ y_R]^T$ and Γ is the matrix containing the cross-probabilities between groups, $\Gamma_{ij} = (\gamma_{j|i})$. Optimal allocations for the three groups can then be found by differentiating the estimated total vote share, $X^{\mathrm{mf}} = \boldsymbol{\rho}^T \boldsymbol{x}$, with respect to the allocation parameters, $\boldsymbol{w}_a$, and equating to zero: $\nabla_{\boldsymbol{w}_a} X^{\mathrm{mf}} = 0$, leading to a system of three non-linear, polynomial equations that can be solved numerically.

To test the accuracy of the mean–field approximation, we compare predictions for stationary vote shares for the three groups of nodes to exact analytical solutions based on Eq. (3.2), for which we take the full network structure into account. To devise a set of test scenarios, we assume that the B-controller distributes her resources equally among different fractions ρ_B of nodes of the network, while we spread allocations from the A-controller uniformly across the network (i.e. $w_{ai} = \mathcal{B}_a/N\ \forall i$).

Figure A.6 shows the differences in vote share between exact, $x_{\mathbf{S}}$, and approximate, $x_{\mathbf{S}}^{\mathrm{mf}}$, estimations of vote shares for each of the three groups, and for scenarios with a varying fraction of B-targeted nodes (ρ_B) and B-controller's budget ($\mathcal{B}_b$). As expected, we generally observe a decline in vote shares with increasing $\mathcal{B}_b$. Qualitative dependencies are well captured by the mean–field approach. In more detail, the left panel of Fig. A.6 shows the case of $\rho_B = 0.01$, implying that roughly a fraction $\rho_N = 0.03$ of nodes are adjacent to them. We note that the mean vote share of nodes $\mathbf{T}_b$

and **R** is accurately estimated by the mean–field approximation ($x_B^{\mathrm{mf}} \approx x_B, x_R^{\mathrm{mf}} \approx x_R$), while the vote share of the nodes in $\mathbf{N}_b$ is slightly overestimated over the whole range of budgets ($x_N^{\mathrm{mf}} > x_N$). In the middle panel, we set $\rho_B = 0.1$, implying $\rho_N = 0.24$ and $\rho_R = 0.66$. For this scenario, mean–field results for all groups are in very good agreement with numerical results. Last, as shown in the right hand panel of Fig. A.6, we also consider $\rho_B = 0.3$, $\rho_N = 0.46$, $\rho_R = 0.24$. Mean–field estimates are again in good agreement, but some slight systematic underestimation of x_R is observed.

In general, mean–field estimates for nodes in $\mathbf{T}_b$ prove highly accurate. This is expected, as nodes in $\mathbf{T}_b$ are heavily influenced by the B-controller and less susceptible to network effects. The estimation errors for the other two groups are caused by two different limitations of the approximations. The first limitation results from treating all nodes in **R** as a uniform group, even though they have different distances to nodes in $\mathbf{N}_b$ and, consequently, different vote shares. This limitation is most prominent for low densities of ρ_B and we see its effects in Fig. A.6–left. Here, we observe an over-estimation of x_N, which is caused by an over-estimation of vote shares of second neighbours of B-controlled nodes. The second limitation of our mean–field approach is not subdividing nodes in $\mathbf{N}_b$ into separate classes depending on exact numbers of adjacent nodes in $\mathbf{T}_b$ or **R**. However, as nodes in $\mathbf{T}_b$ are allocated randomly, nodes in $\mathbf{N}_b$ may have different numbers of neighbours in $\mathbf{T}_b$. As a result, nodes in $\mathbf{N}_b$ having more neighbours in $\mathbf{T}_b$ than the average will have smaller probabilities of adopting A, but at the same time, a lack of remaining connections to nodes in **R** implies that they have a reduced impact on this last group. In contrast, nodes in $\mathbf{N}_b$ with fewer neighbours belonging to $\mathbf{T}_b$ than the average have higher probabilities to adopt A and a larger impact on **R**. Hence, the mean–field approach underestimates vote shares of nodes in **R**. This situation is particularly prominent when ρ_B is relatively large, which corresponds to the cases shown in Fig. A.6–right.

A.5.2 Analytical Results on Shielding and comparison to Numerical Results

In this section, we compare optimal control allocations obtained numerically via gradient ascent and optima calculated from mean–field estimates via solving $\nabla_{w_a} X^{\mathrm{mf}} = 0$, which allows us to re-assess the roles of shielding and shadowing. As in the previous subsection, experiments are run on K-regular networks with a passive opponent targeting a varying fraction of nodes ρ_B and with varying relative budget $\mathcal{B}_b/\mathcal{B}_a$.

Figure A.7 presents the optimal allocations given by both numerical and analytical solutions (top row) and the intensity of shielding in the allocations (bottom row). In the top row, we note a particular pattern in optimal allocations to nodes in $\mathbf{T}_b$ (in *blue*): allocations are roughly inversely proportional to the budget of the B-controller and

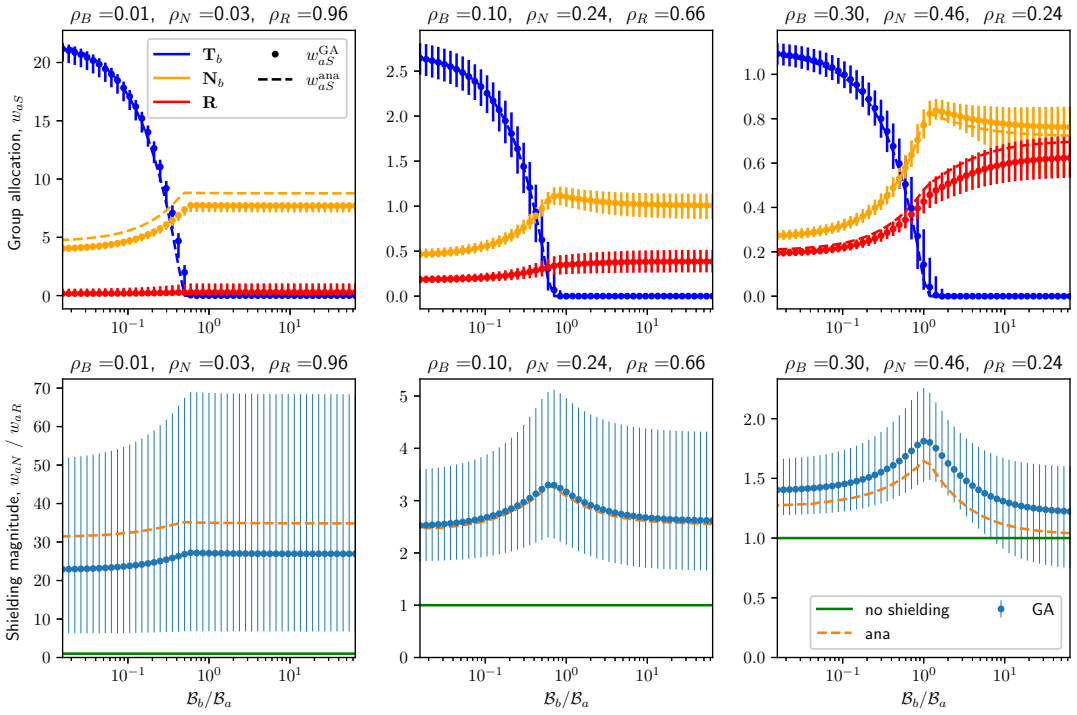


FIGURE A.7: Dependence of optimal control allocation for the three groups in **S** (top) and shielding strength (bottom) on relative budgets ($\mathcal{B}_b/\mathcal{B}_a$) for varying fractions of B-targeted nodes (ρ_B) on 3-regular graphs. Analytical solutions correspond to w_{aS}^{ana} (*dashed lines*), and numerical solutions w_{aS}^{GA} (*dots*) are averaged over 15 randomly generated networks of size $N = 1000$, with vertical lines corresponding to the upper and lower mean absolute deviations of allocations within a group. Titles of panels also show the derived fractions of nodes in $\mathbf{N}_b$ (ρ_N) and nodes in $\mathbf{R}$ (ρ_R). The active controller has a fixed budget of $\mathcal{B}_a = N/2$

vanish at a point near budget equality. This finding relates again to the shadowing behaviour already shown in Sect. 3.5. Regarding nodes in $\mathbf{N}_b$ (in *orange*) and in $\mathbf{R}$ (in *red*), we note that generally $w_{aN} > w_{aR}$, i.e. we establish the presence of a shielding effect for all parameter settings.

In the scenario of $\rho_B = 0.01$ (left), analytical solutions generally assign more budget to nodes in $\mathbf{N}_b$ than numerical solutions. This difference is likely due to the mean–field approximation ignoring the role of second neighbours, while these nodes can be used to produce a second barrier of shielding that protects further neighbours (as was seen in Fig. A.5). The optimal allocations given by both methods match very well for $\rho_B = 0.1$ (middle). For $\rho_B = 0.3$ (right), the analytical allocations underestimate the benefits of shielding. This can be an effect of the mean–field approximation not differentiating between the number of neighbours from $\mathbf{R}$ that a node in $\mathbf{N}_b$ has; only a subgroup of nodes in $\mathbf{N}_b$ is in direct contact with any neighbours in $\mathbf{R}$, and focusing on those makes shielding more effective. For high values of the opponent's budget, there are nodes in $\mathbf{N}_b$ in the numerical solutions that receive even less allocation than nodes in $\mathbf{R}$ (panel bottom–right). These are those nodes in $\mathbf{N}_b$ whose three neighbours

belong to $\mathbf{T}_b$ (not shown), so resources allocated to them do not spread their impact to other nodes and therefore have little effect. The intensity of shielding is higher the lower the ρ_B (panel bottom-left).

A.6 Dependence of optimal influence allocations in the continuous regime on node degree

In this section, we directly show the relation between optimal allocations and node degree in the continuous regime. Fig A.8**a** shows the dependence of optimal allocations on the weighted degree of nodes for the reference case of $K = 16$ random nodes targeted by the opponent. This scenario shows some degree of correlation between optimal influence allocations and node degree (with a Kendall rank coefficient of $\tau = 0.37$, as can be seen in Fig 3.6b). In contrast, Fig A.8**b** shows a similar scenario, but with the passive opponent not choosing her targets randomly but only targeting nodes whose neighbours have the lowest possible degree. We can see here that correlations are not as evident; the Kendall rank coefficient is $\tau = 0.05$, as can be seen in Fig 3.6c. For instance, nodes whose weighted degree is above $d_i = 40$ receive very little allocation, and nodes with degree two, three and four receive higher allocation than most nodes with higher degree.

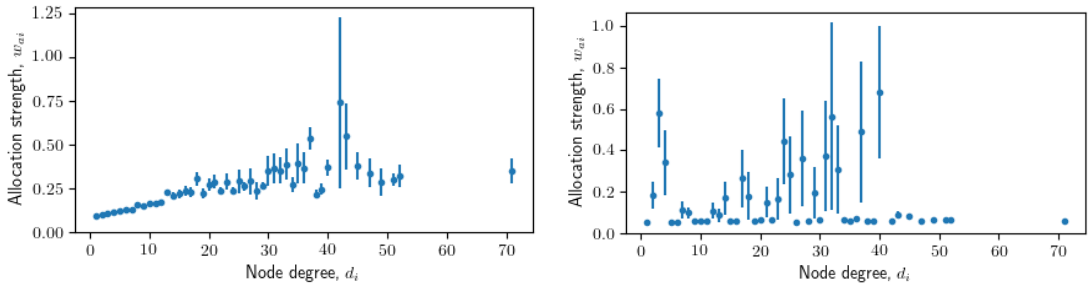


FIGURE A.8: Dependence of optimal influence allocations w_{ai} in the continuous regime on node degree d_i. The passive controller targets $K = 16$ nodes in the network **a** randomly chosen or **b** whose neighbours have the lowest possible degree. Error bars represent standard errors of the mean (over all nodes with same degree). The total budget of both controllers sums up to $\mathcal{B}_a + \mathcal{B}_b = N\langle d \rangle / 30$.

A.7 Shadowing, shielding and hub preferences in the discrete regime

In this section, we investigate the role that shadowing, shielding and node degree have in optimal allocations in the discrete regime against a passive controller who targets random nodes. We illustrate our results through numerical experiments for

which we set both controllers to target K nodes in the network and vary K and the budget ratio $r = \mathcal{B}_a/\mathcal{B}_b$.

To test for the presence of shadowing, we measure the fraction $\phi = \mathbb{P}(i \in \mathbf{T}_a \mid i \in \mathbf{T}_b)$ of nodes targeted by the passive controller ($\mathbf{T}_b = \{i \,|\, w_{bi} = g\}$) that are also targeted by the active controller ($\mathbf{T}_a = \{i \,|\, w_{ai} = g\}$). A strong presence of shadowing would be indicated by high values of ϕ, i.e. $\phi > 0.5$. We take as a reference the fraction ϕ_R obtained by an active controller who just targets random nodes in the network. Results strongly depend on the budget ratio of both controllers, as shown in Fig A.9a. For equal budgets (*rhombi*), the active controller targets around 20% of the nodes targeted by the passive controller ($\phi \approx 0.2$) for most K. This is significantly different from the random behaviour for $K < 50$, with $\phi_R \approx 0$. So we can affirm that there is some presence of shadowing in the equal budget scenario, albeit very weak ($\phi = 0.2$). When in budget advantage (*squares*), shadowing is generally present with $\phi > 0.6$ for most K, so shielding is clearly present. In contrast, when in budget disadvantage (*circles*) the controller completely avoids targeting the nodes targeted by the passive controller, as $\phi \approx 0$ for all K. In conclusion, we see that optimal first-order strategies are qualitatively similar in the continuous and discrete regimes: with shadowing when in budget advantage, avoidance when in budget disadvantage, and a weak form of shadowing when in budget equality.

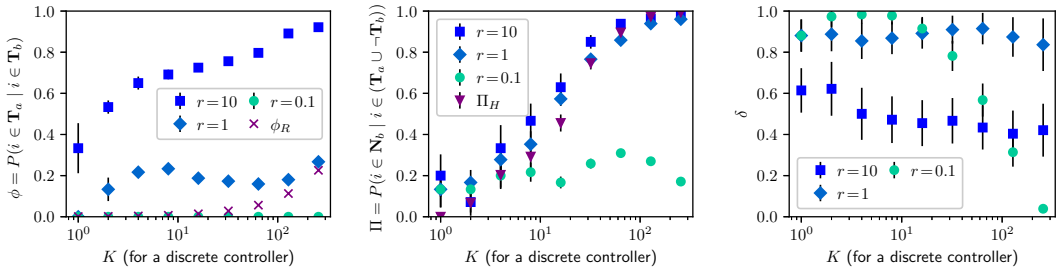


FIGURE A.9: Amount of **a** shadowing, **b** shielding, and **c** degree dependence present in optimal discrete allocations for a range of numbers of nodes K targeted by either controller and different budget ratios $r = \mathcal{B}_a/\mathcal{B}_b$. Error bars represent standard errors over 15 instances of the experiments and the total budget of both controllers sums up to $\mathcal{B}_a + \mathcal{B}_b = N\langle d \rangle/30$.

To test for the presence of shielding in the discrete regime, we measure the probability Π that optimal allocations are directed towards neighbours of the nodes targeted by the passive controller, i.e. we measure $\mathbb{P}(i \in \mathbf{N}_b \mid i \in (\mathbf{T}_a \cap \neg\mathbf{T}_b))$. Shielding would be evident with high values of Π. Note that we decouple the effect of shadowing in the probability by including the condition $i \in \neg\mathbf{T}_b$. Since shielding and node degree are related insofar as nodes with high degree are more likely to be in $\mathbf{N}_b$, we compare the results with what would be obtained by an active controller who targets the nodes with the highest degree in the network, labelled as Π_H. Results are again found to depend on the budget ratio r, as shown in Fig A.9b. When in budget equality (*rhombi*) or superiority (*squares*), shielding-related probabilities Π steadily increase with K.

However, these results are very similar to those obtained by the hub-targeting strategy Π_H so, inversely to what happens in the continuous regime, these values of Π likely are an artefact of a high-degree preference. Different to the other two scenarios, when in budget disadvantage (*circles*) shielding-related probabilities are below $\Pi < 0.4$ and largely unaffected by K. These low values of Π can be linked to an anti-shielding behaviour, where neighbours of the nodes targeted by the opponent tend to be avoided, as shown for the continuous regime (Fig 3.5b).

Last, we investigate the degree dependence of optimal allocations by looking at the mean degree of nodes targeted by the active controller $\langle d \rangle_{\mathbf{T}_a}$. We normalise these measurements via two reference strategies: targeting nodes with the highest (H) and lowest (L) degrees in the network. These reference strategies serve as upper and lower bounds on the obtainable average degree, respectively, with the normalised mean degree as $\delta = (\langle d \rangle_{\mathbf{T}_a} - \langle d \rangle_L)/(\langle d \rangle_H - \langle d \rangle_L)$. Based on the resulting δ, as seen in Fig A.9c, we make three different observations depending on the budget ratio between controllers. First, when in budget equality (*rhombi*), the mean targeted degree stays above $\delta > 0.8$ irrespective of K, so the optimal strategy is greatly focused on targeting high-degree nodes. Second, when in budget disadvantage (*circles*), there is a transition in δ, as $\delta > 0.8$ for $K < 30$ and $\delta < 0.5$ for $K > 100$. This behaviour can be interpreted as an effect of anti-shielding: when K is low, hubs are targeted often since there is a preference towards hubs and they have low chances of belonging to $\mathbf{N}_b$. However, for large K, nodes with high degree almost surely belong to the group $\mathbf{N}_b$, which tend to be avoided, as seen in Fig A.9b. Last, when in budget advantage (*squares*), mean targeted degrees are in the range $\delta \in (0.4, 0.65)$ for all K. The lower δ as compared to the budget equality scenario is a result of the preference towards high-degree nodes being partially discounted by the shadowing behaviour.

To summarise, we have seen the effect that the three proposed heuristics have in the discrete regime. First, related to degree dependence, optimal strategies in the discrete regime strongly favour targeting high-degree nodes in most scenarios. This preference only gets discounted by shadowing when the active controller is in budget advantage, and by anti-shielding when in budget disadvantage if controllers target many nodes in the network. Second, the presence of first-order strategies is only significant when in budget inequality, leading to shadowing when in budget advantage and to avoidance when in budget disadvantage. Shielding is not present in any scenario of the discrete regime, although anti-shielding is present when in budget disadvantage.

A.8 Extension to other network topologies

In this section, we have extended a selection of the experiments in Chapter 3 to other complex network topologies. We have chosen two synthetic networks

—Barabasi–Albert networks (Dezső and Barabási, 2002) and scale-free networks built via configuration models (Bayati et al., 2010) —and two real–world networks —a trust network (Massa et al., 2009) and a friendship network (Mastrandrea et al., 2015). Barabasi–Albert (BA) networks have been generated with each new node linking to five previous nodes. The scale-free networks have been generated via the configuration model (Molloy and Reed, 1995) with an exponent of $\gamma = 3$ and setting the minimum number of connections per node to five. These parameters have been chosen experimentally in such a way as to generate networks with similar connectivity level $\langle d \rangle$ as in the email interaction network from Chapter 3. Similarly, networks have been generated with the same number of nodes as the e-mail interaction network, namely $N = 1133$. Among the real–world networks, the online social network data corresponds to a weighted trust network among the users from the online social network *Advogato* (Massa et al., 2009; Rossi and Ahmed, 2015). The other real–world network corresponds to a self–reported friendship network between students in a high school in Marseilles, France, in December 2013 (Mastrandrea et al., 2015). For both real–world network, we have employed the biggest component of the network, ignored self–loops, and made the networks undirected. Table A.1 contains summary statistics of all studied networks. Note that the Friendship network has a stronger assortativity (0.26) than the other networks (< 0.1).

Network	**N**	**Mean degree**	**Max degree**	**Assortativity**
Email interaction	1133	9.62	71	0.078
Barabasi-Albert	1133	9.96	129.2 ± 4.3	-0.051 ± 0.003
Scale-free	1133	9.22 ± 0.03	110.0 ± 1.0	-0.043 ± 0.003
Advogato	5054	13.13	747	-0.095
Friendship	128	6.27	17	0.260

TABLE A.1: Statistics of the network employed on the main manuscript (Email interaction) and the networks employed here for the extension of experiments (Barabasi-Albert, Scale-free, Advogato, and Friendship).

In more detail, we have repeated the experiments from Fig 3.1b —profile of optimal allocations—, Fig 3.2b —performance enhancement of using continuous allocations over discrete allocations—, Fig 3.5b —shadowing and shielding behaviours in the continuous regime—, Fig 3.6b —correlations between allocation strength and node degree—, and Fig 3.7—comparison of various heuristics to optimal allocations.

Fig A.10 shows the profile of optimal allocation distribution for each network topology. These figures are analogous to Fig3.1b in the main manuscript. For the real–world networks Advogato and Friendship, the number K of nodes targeted by the opponent has been chosen as to preserve the percentage of nodes in the network targeted. Optimal allocations for the five studied network topologies share a very similar profile, with a smoother allocation profile visible for the Friendship network.

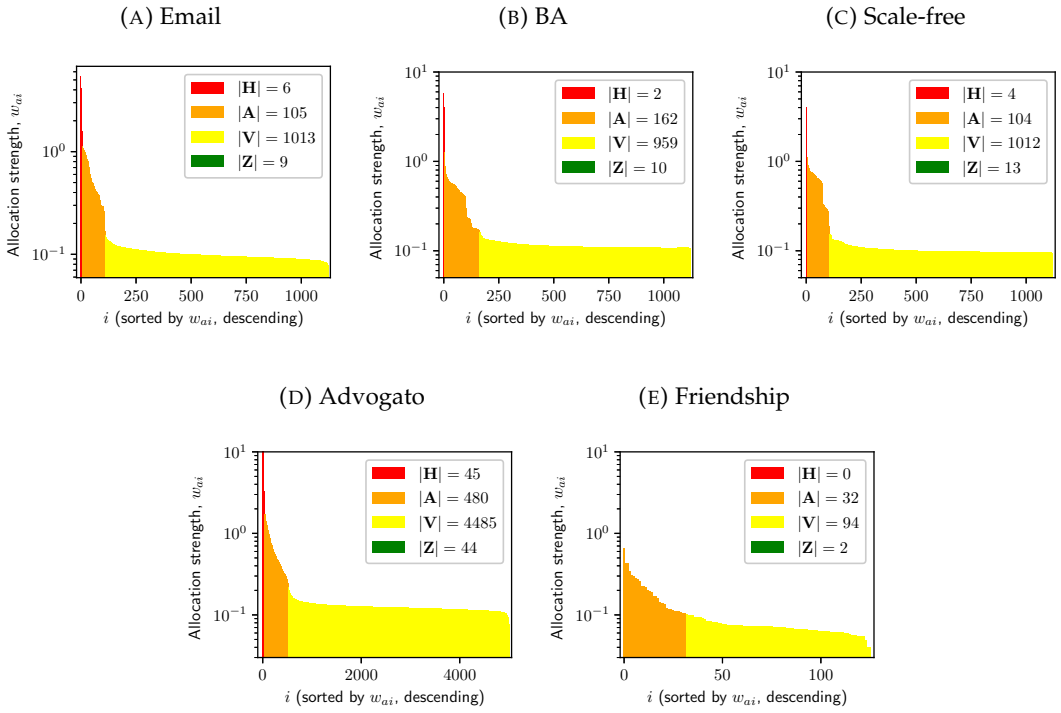


FIGURE A.10: Distribution of optimal influence allocations for various network topologies. Allocations are sorted in descending order and the passive controller discretely targets around 1.5% of the nodes in the network (i.e., **a, b, c** $K = 16$, **d** $K = 71$, **e** $K = 2$), randomly chosen, and both controllers hold the same budget $\mathcal{B} = N\langle d_i \rangle / 60$. Allocations are coloured by the allocation group in $\mathbf{G}$ they belong to.

Fig A.11 shows the percentage enhancement of control of continuous versus discrete optimal allocations for all network topologies studied here. These figures are analogous to Fig 3.2b in the main manuscript. The enhancement of control is generally similar for the five studied network topologies. A stronger effect can be seen in the Friendship network, where enhancement go up to 25% when budgets are equal and the opponent targets only one node in the network.

Fig A.12 shows the average allocation strength given to nodes in the allocation groups related shadowing and shielding, $\mathbf{S} = \{\mathbf{T}_b, \mathbf{N}_b, \mathbf{R}\}$ for all network topologies studied. These figures are analogous to Fig 3.5b in the main manuscript. Shadowing and shielding generally have a similar effect in all of the studied networks, although shadowing is more and less present in the Advogato and Friendship networks, respectively. This effect could be related to the network size.

Fig A.13 compares the Kendall rank correlation coefficients τ between allocation strength w_{ai} and node degree d_i between optimal allocations (*squares/triangles*) and allocations purely driven by a shielding heuristics (*circles*). These figures are analogous to Fig 3.6b in the main manuscript. The five network topologies show very similar behaviour regarding these correlations, both in optimal allocations and shielding heuristics.

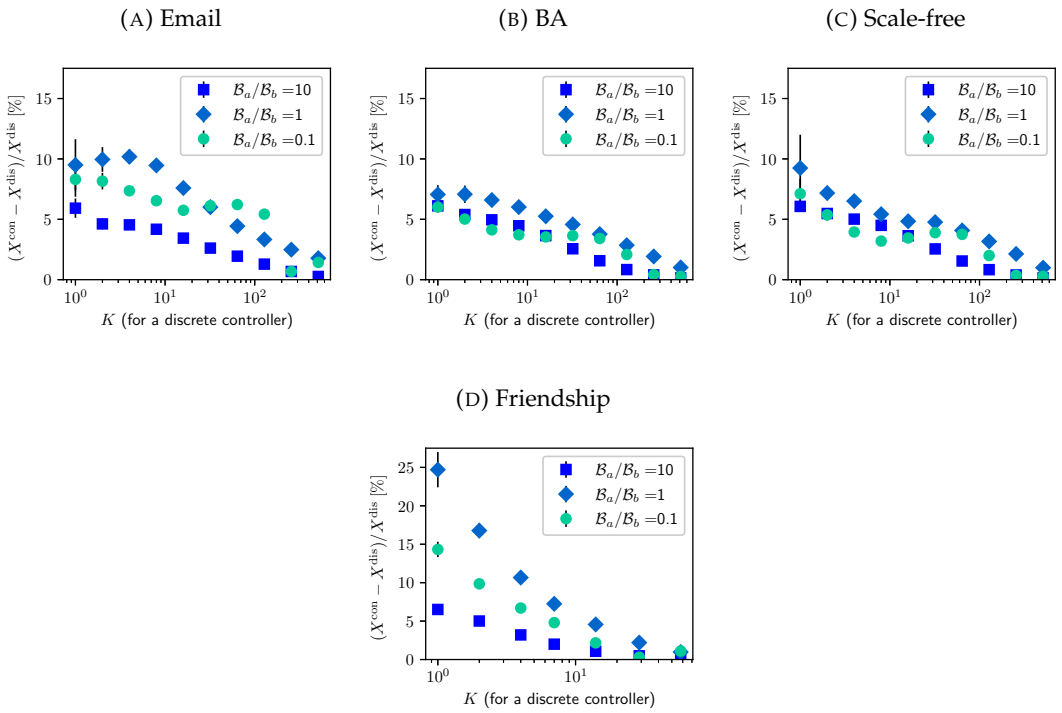


FIGURE A.11: Percentage enhancement of control by optimal continuous allocations over optimal discrete allocations for various network topologies, with varying K (number of targeted nodes by discrete strategies) and budget ratios. Error bars represent standard errors over 15 experiment samples (5 in the case of **d**) and the total budget of both controllers sums up to $\mathcal{B}_a + \mathcal{B}_b = N\langle d\rangle/30$.

Last, Figs A.14 and A.15 explore the effectiveness of different heuristics as compared to the optimised numerical results in the network topologies studied. The performance of each heuristic is captured by the difference in vote share ΔX between the heuristic and the numerically optimised numerical result. The smaller the gap in obtainable vote share ΔX between a heuristic and the numerically determined optimum, the stronger its contribution of the heuristic to optimal allocations. The figures illustrate results of the vote share gap for the five network topologies studied here, for three budget scenarios (advantage, equality, and disadvantage), and against a discrete passive controller who targets around 1.5% of the nodes in the network randomly chosen. Note that we make a clear separation between the discrete (*purple*) and continuous (*orange*) regimes and that their vote shares are compared to the discrete or continuous optimal strategies, respectively. The details of the implementation of each heuristic can be found in Sect. 3.3.3. These figures are analogous to Fig 3.7 in the main manuscript. Heuristics perform very similarly on four of the five network topologies in all scenarios. However, the network Friendship from Fig A.15 shows a remarkably different behaviour in the discrete regime in that the degree–dependent heuristic does not perform significantly better than the other discrete heuristics.

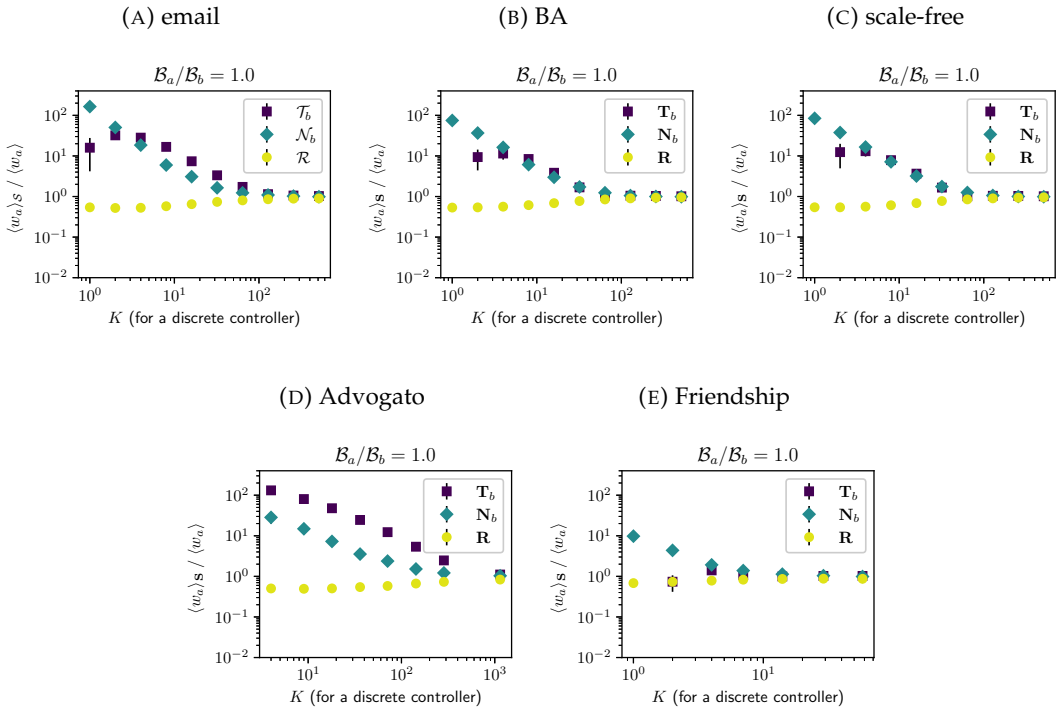


FIGURE A.12: Average optimal allocations given to groups related to shadowing and shielding in various network topologies. Average allocations are normalised with respect to the total average allocation $\langle w_{ai} \rangle$ and both controllers hold equal budgets $\mathcal{B} = N\langle d \rangle / 60$.. $\mathbf{T}_b$ is the set of nodes targeted by the passive opponent, $\mathbf{N}_b$ is the set of nodes that are neighbours of the nodes in $\mathbf{T}_b$, and $\mathbf{R}$ comprises the remaining nodes that do not belong to any of the other two groups. Error bars represent standard errors over 15 experiment samples (5 in the case of $\mathbf{d}$).

A.9 Proof of concavity

We want to prove that the goal function $X = 1/N\ \mathbf{1}^T (L + W_a + W_b)^{-1} W_a \mathbf{1}$ is concave with respect to $\boldsymbol{w_a} \in \mathbb{R}^+$. This is important as it guarantees that the gradient ascent optimisation technique reaches the global maximum of the problem.

We note that our model can be mapped to that of Abebe et al. (2018), which considers opinion dynamics with innate opinions $s_i \in [0, 1]$ and different susceptability of persuasion $\alpha_i \in [0, 1]$ for each node, given by

$$x_i(t+1) = \alpha_i s_i + (1 - \alpha_i) \frac{\sum_{j \in N(i)} x_j(t)}{deg(i)}\ , \tag{A.7}$$

where $N(i)$ is the neighbourhood of node i and $deg(i)$ is the degree of node i (d_i). This expression of the dynamics can be mapped to ours (Eq. 3.1 of the main manuscript) by applying the transformations $s_i = w_{ai}/(w_{ai} + w_{bi})$ and
$\alpha_i = (w_{ai} + w_{bi})/(d_i + w_{ai} + w_{bi})$. Similar to our work, they maximise the vote share of opinions in the equilibrium by optimally adjusting α_i and prove the concavity of the

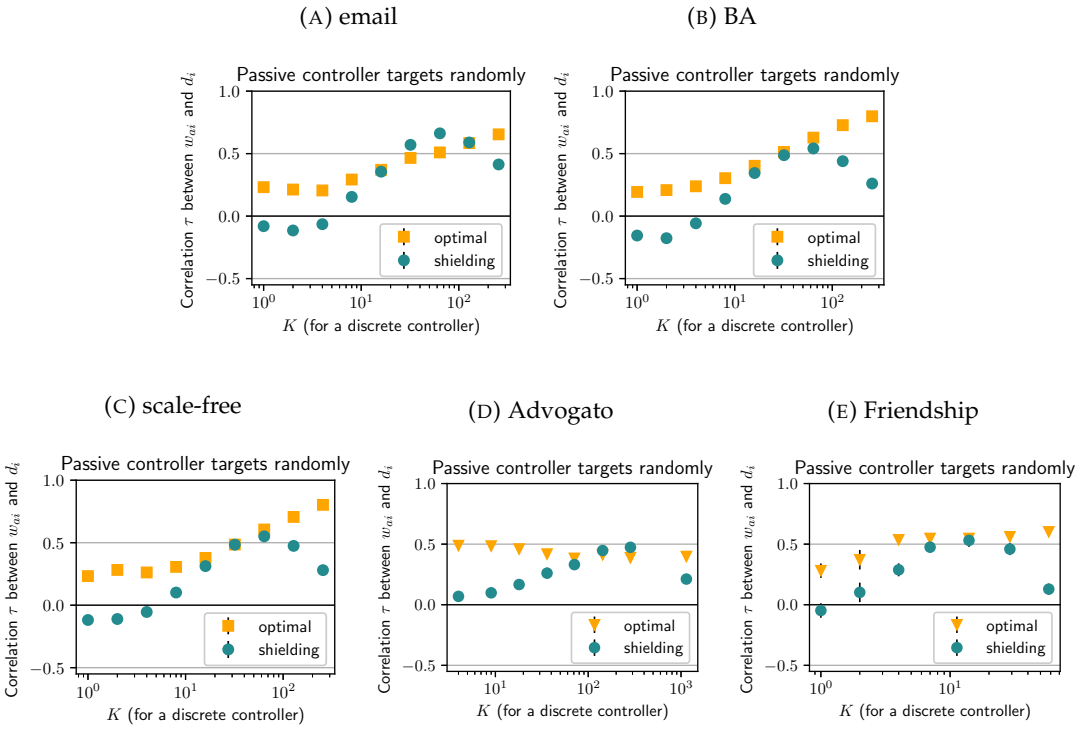


FIGURE A.13: Kendall rank correlation coefficients τ_{wd} between allocation strength w_{ai} and node degree d_i in the continuous and discrete regimes and for various network topologies. The active controller targets nodes optimally (*squares*) or following a shielding strategy (*circles*). The passive controller targets random nodes in the network and the total budget of both controllers sums up to $\mathcal{B}_a + \mathcal{B}_b = N\langle d\rangle/30$. Error bars represent standard errors over 15 experiment samples (5 in the case of **d**).

goal function, $Z = \mathbf{1}^T[\,I - (I-A)P\,]^{-1}A\boldsymbol{s}$, with respect to α_i, where A is a diagonal matrix with $A_{ii} = \alpha_i$, P is the random walk matrix $P = D^{-1}W$, and D is a diagonal matrix with $D_{ii} = d_i$. But Z is also equivalent to our vote share X, as

$$\begin{aligned} Z = \mathbf{1}^T[\,I - (I-A)P\,]^{-1}A\boldsymbol{s} &\overset{(A.9)}{=} \\ = \mathbf{1}^T[I - [I - (W_a+W_b)(D+W_a+W_b)^{-1}]\,D^{-1}W]^{-1}(D+W_a+W_b)^{-1}W_a\mathbf{1} &\overset{(A.10)}{=} \\ = \mathbf{1}^T(D+W_a+W_b-W)^{-1}W_a\mathbf{1} &\overset{(A.11)}{=} \\ = \mathbf{1}^T(L+W_a+W_b)^{-1}W_a\mathbf{1} &\overset{(3.3)}{=} N\,X\,, \end{aligned} \tag{A.8}$$

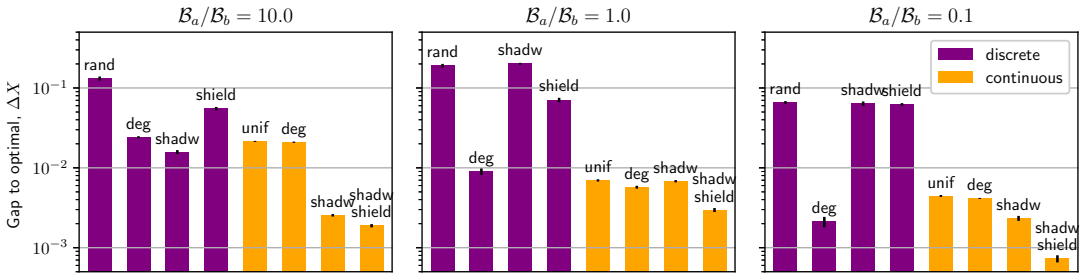
where we have used the following facts:

$$A = (W_a+W_b)(D+W_a+W_b)^{-1}, \qquad \boldsymbol{s} = W_a(W_a+W_b)^{-1}\mathbf{1}\,, \tag{A.9}$$

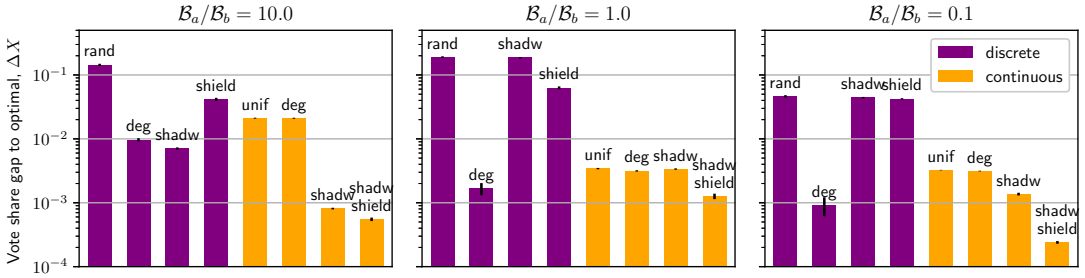
$$O^{-1}Q^{-1} = (O\,Q)^{-1} \quad \text{(provided both matrices are symmetric)}, \tag{A.10}$$

$$L = D - W. \tag{A.11}$$

(A) email



(B) BA



(C) scale-free

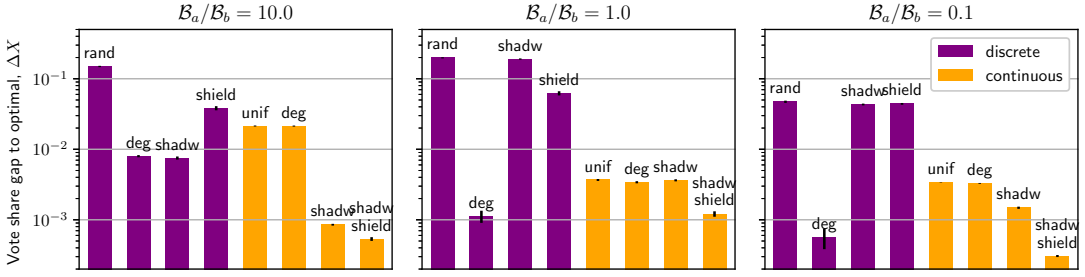


FIGURE A.14: Comparison of various heuristics to optimal allocations. Bars represent the gap in vote share ΔX of the heuristics with respect to optimal numerical allocations for three different budget scenarios (as indicated on the top of the panels). Each bar represents one of the following heuristics: random (*rand*), degree-based (*deg*), shadowing-based (*shadw*), shielding-based (*shield*), uniform targeting (*unif*), combination of shadowing and shielding (*shadw shield*). The passive controller targets $K = 16$ random nodes in the network in a discrete fashion. Error bars represent standard errors for 15 instances of the experiments and the total budget of both controllers sums up to $\mathcal{B}_a + \mathcal{B}_b = N\langle d \rangle / 30$.

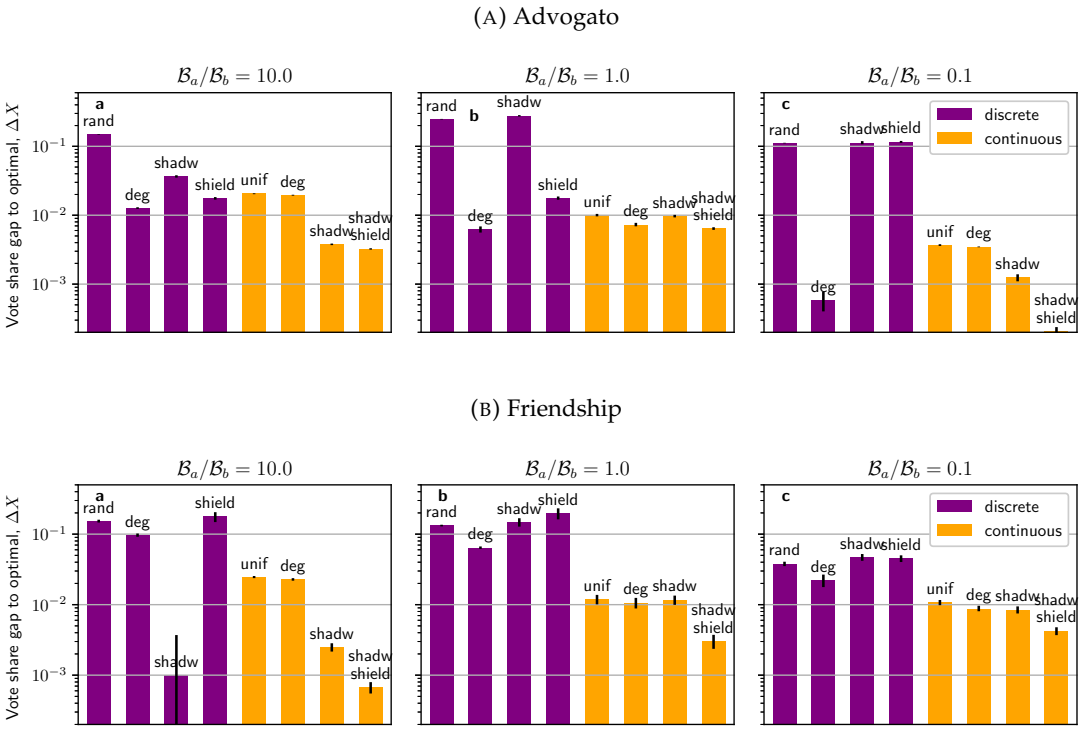


FIGURE A.15: Comparison of various heuristics to optimal allocations. Bars represent the gap in vote share ΔX of the heuristics with respect to optimal numerical allocations for three different budget scenarios (as indicated on the top of the panels). Each bar represents one of the following heuristics: random (*rand*), degree-based (*deg*), shadowing-based (*shadw*), shielding-based (*shield*), uniform targeting (*unif*), combination of shadowing and shielding (*shadw shield*). The passive controller targets **a** $K = 71$ or **b** $K = 2$ random nodes in the network in a discrete fashion. Error bars represent standard errors for 15 instances of the experiments and the total budget of both controllers sums up to $\mathcal{B}_a + \mathcal{B}_b = N\langle d \rangle / 30$.

Since the mappings in (A.9) are monotonic in W_a, the proof of concavity from Abebe et al. (2018) can be applied to our problem by following analogous steps.

A.10 Iterations to optimal IM of the proposed heuristics

This section explores an alternative assessment of the proposed heuristics for IM. In Fig 3.7 of the main manuscript, we investigate gaps in vote share ΔX between each of the heuristics and the optimal strategy obtained numerically. Here, we provide an alternative assessment of the heuristics measuring computational costs to achieve similar quality solutions by performing numerical optimisation using gradient ascent. To achieve this we measure how many iterations of gradient-ascent are needed to reach vote shares X corresponding to each of the heuristics (Fig A.16). That is, good heuristics correspond to large a number of iterations (i.e. high computational costs to achieve similar solutions) and poor heuristics to low numbers of iterations (or low

computational costs to achieve similar outcomes by optimisation). From the figure, we can observe that all discrete strategies (*purple*) are already surpassed with less than 10 iterations of the gradient-ascent algorithm. In fact, in most cases the discrete *shadowing* and *shielding* strategies are already surpassed with a single iteration of the algorithm. Continuous heuristics (*orange*) are generally one order of magnitude better than discrete strategies (*purple*).

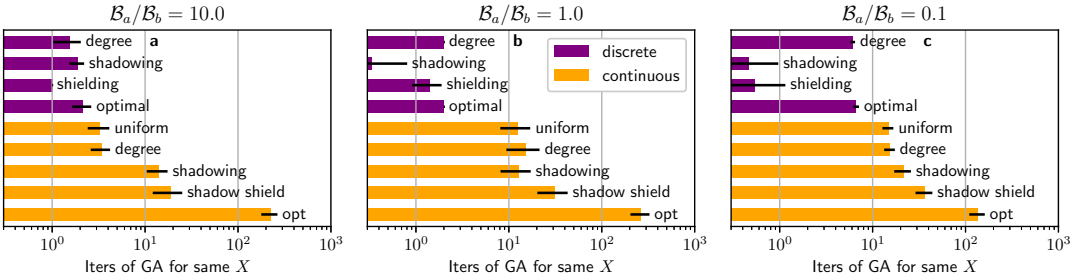


FIGURE A.16: Alternative assessment of the proposed heuristics. On the x-axis, the number of iterations of the gradient-ascent algorithm required for obtaining a similar level of performance than the corresponding heuristic. On the y-axis, bars representing one of the following heuristics: degree-based (*degree*), shadowing-based (*shadowing*), shielding-based (*shielding*), uniform targeting (*uniform*), combination of shadowing and shielding (*shadow shield*). The passive controller randomly targets $K = 16$ nodes in the network in a discrete fashion and the total budget of both controllers sums up to $\mathcal{B}_a + \mathcal{B}_b = N\langle d \rangle/30$. Three budget scenarios are depicted (as indicated on the top of the panels). Error bars represent standard errors for 15 instances of the experiments. The initial allocation of the gradient-ascent algorithm is set at a discrete allocation on K random nodes.

Appendix B

Appendix to Chapter 4

B.1 Phase diagram of optimal targetting for an active controller who is in budget superiority against a passive controller with varying targetting strategies

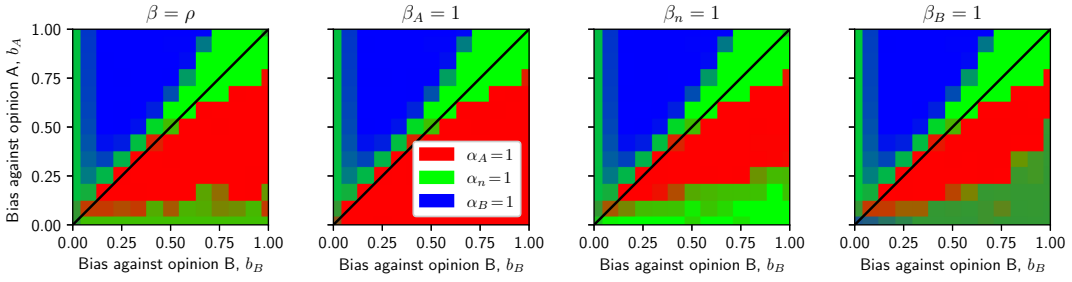


FIGURE B.1: Target group of optimal allocations by the active controller when in budget superiority ($\hat{\mathcal{B}}_a = 10\hat{\mathcal{B}}_b$), for different combinations for the levels of bias b_A and b_A, when the opponent targets all agents equally ($\beta = \rho$) or focuses all her influence on agents biased against the active controller ($\beta_A = 1$), normal agents ($\beta_n = 1$), or agents biased against her opinion ($\beta_B = 1$), on a complete graph of size $N = 100$.

Appendix C

Appendix to Chapter 5

C.1 Role of system size on consensus–forming process under party competition

Here, we explore the role of system size — i.e. number of agents N — on the effects that party competition has on opinion dynamics outcomes. Fig. C.1 presents similar experiments as in Fig. 5.1**b**, but varying N over several orders of magnitude. From the left panel, with a system size 100 times smaller than in Chapter 5, it can be seen that low number of agents introduces substantial stochasticity in the model and the phase transition structure is mostly lost. However, the consensus promoting effect can still be perceived, as the effective number of clusters is consistently lower when party competition is present. The other panels, starting from $N = 10^4$ and with system sizes increasing by factors of 10, reveal a phase landscape with increasing sharpness, as the widths of the phase transitions narrow. It could be expected that in the limit of $N \to \infty$, phase transitions become step functions. In light of these observations, we can conclude that the effect of finite sizes does not qualitatively affect the results shown in Chapter C with populations larger than $N \geq 10^4$.

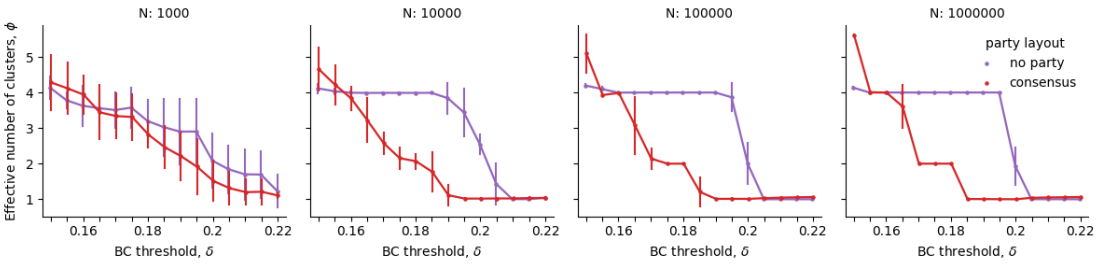


FIGURE C.1: Effects of different system sizes N: 10^3 (left-most), 10^4 (middle–left), 10^5 (middle–right), 10^6 (right-most). Points represent mean values and error bars standard deviations over 10 runs of the experiment.

C.2 Phase diagrams of transitions from consensus to polarisation to fragmentation for different number K of parties

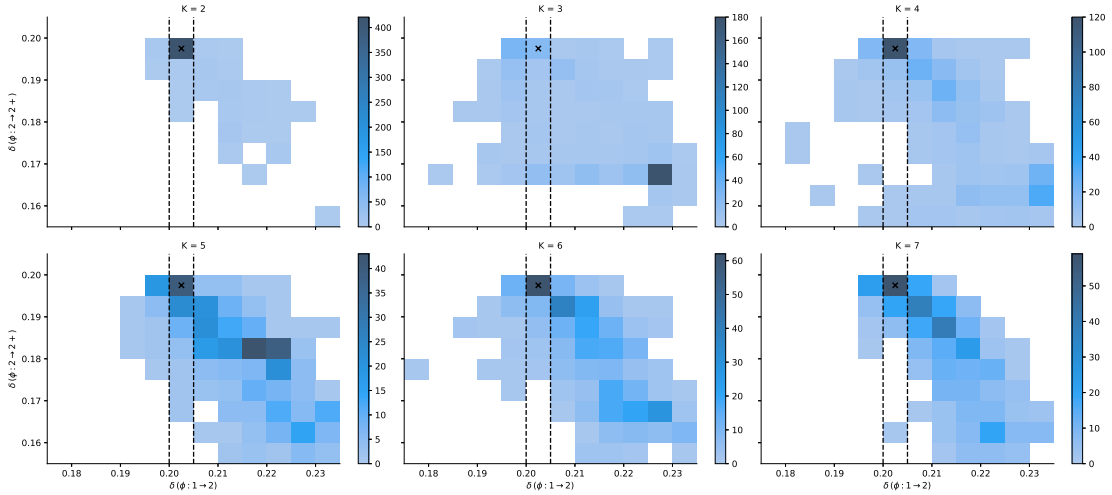


FIGURE C.2: Histogram of phase transitions $\delta_{\phi:1\to2}$ and $\delta_{\phi:2\to2+}$ for 500 randomly sampled party configurations for each $K \in \{2,\ldots,7\}$. The black cross indicates the phase transitions of the no–party baseline, while solid lines separate the different classes of outcomes, from left to right: fostering consensus, no effect on the opinion dynamics, and fostering polarisation.

References

Abebe, R., Kleinberg, J., Parkes, D., and Tsourakakis, C. E. (2018). Opinion Dynamics with Varying Susceptibility to Persuasion. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '18*, pages 1089–1098, New York, New York, USA. ACM Press.

Abramowitz, A. I. and Saunders, K. L. (2008). Is Polarization a Myth? *The Journal of Politics*, 70(2):542–555.

Adams, J. and Merril, S. (1999). Modeling Party Strategies and Policy Representation in Multiparty Elections: Why Are Strategies so Extreme? *American Journal of Political Science*, 43(3):765.

Albanese, F., Pinto, S., Semeshenko, V., and Balenzuela, P. (2020). Analyzing mass media influence using natural language processing and time series analysis. *Journal of Physics: Complexity*, 1(2):025005.

Albi, G., Herty, M., and Pareschi, L. (2015). Kinetic description of optimal control problems and applications to opinion consensus. *Communications in Mathematical Sciences*, 13(6):1407–1429.

Ali, K., Wang, C.-Y., Yeh, M.-Y., and Chen, Y.-S. (2020). Addressing Competitive Influence Maximization on Unknown Social Network with Deep Reinforcement Learning. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, volume 1, pages 196–203. IEEE.

Alshamsi, A., Pinheiro, F. L., and Hidalgo, C. A. (2018a). Optimal diversification strategies in the networks of related products and of related research areas. *Nature Communications*, 9(1):1328.

Alshamsi, A., Pinheiro, F. L., and Hidalgo, C. A. (2018b). Optimal diversification strategies in the networks of related products and of related research areas. *Nature Communications*, 9(1):1328.

Altarelli, F., Braunstein, A., Dall'Asta, L., Wakeling, J. R., and Zecchina, R. (2014). Containing epidemic outbreaks by message-passing techniques. *Physical Review X*.

Alvarez, R. M. and Nagler, J. (1998). When Politics and Models Collide : Estimating Models of Multiparty Elections. *American Journal of Political Science*, 42(1):55–96.

Alvarez, R. M., Nagler, J., and Bowler, S. (2000). Issues, Economics, and the Dynamics of Multiparty Elections: The British 1987 General Election. *American Political Science Review*, 94(1):131–149.

Amblard, F., Bouadjio-Boulic, A., Sureda Gutiérrez, C., and Gaudou, B. (2015). Which models are used in social simulation to generate social networks? a review of 17 years of publications in JASSS. In *2015 Wint. Sim. Conf.*, pages 4021–4032.

Amorós, P. and Puy, M. S. (2010). Indicators of electoral victory. *Public Choice*, 144(1-2):239–251.

Anderson, S. P., de Palma, A., and Thisse, J.-F. (1992). *Discrete Choice Theory of Product Differentiation*. The MIT Press, Cambridge, MA.

Aral, S. and Dhillon, P. S. (2018). Social influence maximization under empirical influence models. *Nature Human Behaviour*, 2(6):375–382.

Arendt, D. L. and Blaha, L. M. (2015). Opinions, influence, and zealotry: a computational study on stubbornness. *Computational and Mathematical Organization Theory*, 21(2):184–209.

Atienza-Barthelemy, J., Martin-Gutierrez, S., Losada, J. C., and Benito, R. M. (2019). Relationship between ideology and language in the Catalan independence context. *Scientific Reports*, 9(1):17148.

Axelrod, R. (1997a). *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton University Press.

Axelrod, R. (1997b). The Dissemination of Culture. *Journal of Conflict Resolution*, 41(2):203–226.

Axelrod, R., Daymude, J. J., and Forrest, S. (2021). Preventing extreme polarization of political attitudes. *Proceedings of the National Academy of Sciences*, 118(50).

Axelrod, R. and Hamilton, W. D. (1981). The evolution of cooperation. *science*, 211(4489):1390–1396.

Badawy, A., Ferrara, E., and Lerman, K. (2018). Analyzing the digital traces of political manipulation: The 2016 russian interference twitter campaign. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 258–265.

Bakshy, E., Hofman, J. M., Mason, W. A., and Watts, D. J. (2011). Everyone's an influencer: Quantifying influence on twitter. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, page 65–74, New York, NY, USA. Association for Computing Machinery.

Bakshy, E., Rosenn, I., Marlow, C., and Adamic, L. (2012). The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web - WWW '12*, pages 519–528, New York, New York, USA. ACM Press.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2013). The Diffusion of Microfinance. *Science*, 341(6144):1236498.

Barabási, A.-L. and Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512.

Battiston, F., Cencetti, G., Iacopini, I., Latora, V., Lucas, M., Patania, A., Young, J.-G., and Petri, G. (2020). Networks beyond pairwise interactions: Structure and dynamics. *Physics Reports*, 874(June):1–92.

Bayati, M., Kim, J. H., and Saberi, A. (2010). A Sequential Algorithm for Generating Random Graphs. *Algorithmica*, 58(4):860–910.

Bharathi, S., Kempe, D., and Salek, M. (2007). Competitive Influence Maximization in Social Networks. In *Internet and Network Economics*, pages 306–311. Springer Berlin Heidelberg, Berlin, Heidelberg.

Bhat, D. and Redner, S. (2019). Nonuniversal opinion dynamics driven by opposing external influences. *Physical Review E*, 100(5):050301.

Bhat, D. and Redner, S. (2020). Polarization and consensus by opposing external sources. *Journal of Statistical Mechanics: Theory and Experiment*, 2020(1):013402.

Bimpikis, K., Ozdaglar, A., and Yildiz, E. (2016). Competitive Targeted Advertising Over Networks. *Operations Research*, 64(3):705–720.

Bischof, D. and Wagner, M. (2019). Do voters polarize when radical parties enter parliament? *American Journal of Political Science*, 63(4):888–904.

Biswas, S., Lima, F. W. S., and Flomenbom, O. (2018). Are Socio-Econo-Physical Models Better to Explain Biases in Societies? *Reports in Advances of Physical Sciences*, 02(02):1850006.

Borondo, J., Morales, A. J., Losada, J. C., and Benito, R. M. (2012). Characterizing and modeling an electoral campaign in the context of Twitter: 2011 Spanish Presidential election as a case study. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(2):023138.

Boudin, L., Monaco, R., and Salvarani, F. (2010). Kinetic model for multidimensional opinion formation. *Physical Review E*, 81(3):036109.

Boxell, L., Gentzkow, M., and Shapiro, J. M. (2017). Is the internet causing political polarization? Evidence from demographics. Technical report, National Bureau of Economic Research.

Braha, D. and de Aguiar, M. A. M. (2017). Voting contagion: Modeling and analysis of a century of U.S. presidential elections. *PLOS ONE*, 12(5):e0177970.

Brede, M., Restocchi, V., and Stein, S. (2018). Resisting influence: How the strength of predispositions to resist control can change strategies for optimal opinion control in the voter model. *Frontiers in Robotics and AI*, 5:34.

Brede, M., Restocchi, V., and Stein, S. (2019a). Effects of time horizons on influence maximization in the voter dynamics. *Journal of Complex Networks*, 7(3):445–468.

Brede, M., Restocchi, V., and Stein, S. (2019b). Transmission errors and influence maximization in the voter model. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(3):033401.

Brede, M. and Romero-Moreno, G. (2022). Sensing Enhancement on Social Networks: The Role of Network Topology. *Entropy*, 24(5):738.

Brede, M. and Romero Moreno, G. (2022). Sensing enhancement on complex networks. In Benito, R. M., Cherifi, C., Cherifi, H., Moro, E., Rocha, L. M., and Sales-Pardo, M., editors, *Complex Networks & Their Applications X*, pages 353–364, Cham. Springer International Publishing.

Broido, A. D. and Clauset, A. (2019). Scale-free networks are rare. *Nature Communications*, 10(1):1017.

Brooks, H. Z. and Porter, M. A. (2020). A model for the influence of media on the ideology of content in online social networks. *Physical Review Research*, 2(2):023041.

Budak, C., Agrawal, D., and El Abbadi, A. (2011). Limiting the spread of misinformation in social networks. In *Proceedings of the 20th international conference on World wide web - WWW '11*, page 665, New York, New York, USA. ACM Press.

Budge, I. and Farlie, D. (1983). *Explaining and predicting elections : issue effects and party strategies in twenty-three democracies*. Allen & Unwin.

Burden, B. C. (1997). Deterministic and Probabilistic Voting Models. *American Journal of Political Science*, 41(4):1150.

Cai, Z., Brede, M., and Gerding, E. (2021). Influence Maximization for Dynamic Allocation in Voter Dynamics. In *Studies in Computational Intelligence*, volume 943, pages 382–394. Springer International Publishing.

Cai, Z., Gerding, E., and Brede, M. (2022). Accelerating opponent strategy inference for voting dynamics on complex networks. In Benito, R. M., Cherifi, C., Cherifi, H., Moro, E., Rocha, L. M., and Sales-Pardo, M., editors, *Complex Networks & Their Applications X*, pages 844–856, Cham. Springer International Publishing.

Campbell, J. E. (2018). *Polarized: Making sense of a divided America*. Princeton University Press.

Carletti, T., Fanelli, D., Grolli, S., and Guarino, A. (2006). How to make an efficient propaganda. *Europhysics Letters (EPL)*, 74(2):222–228.

Carothers, T. and O'Donohue, A. (2019). *Democracies divided: The global challenge of political polarization*. Brookings Institution Press.

Carro, A., Toral, R., and San Miguel, M. (2013). The Role of Noise and Initial Conditions in the Asymptotic Solution of a Bounded Confidence, Continuous-Opinion Model. *Journal of Statistical Physics*, 151(1-2):131–149.

Carro, A., Toral, R., and San Miguel, M. (2016). The noisy voter model on complex networks. *Scientific Reports*, 6(1):24775.

Castellano, C., Fortunato, S., and Loreto, V. (2009a). Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2):591–646.

Castellano, C., Muñoz, M. A., and Pastor-Satorras, R. (2009b). Nonlinear q-voter model. *Physical Review E*, 80(4):041129.

Chakraborty, S., Stein, S., Brede, M., Swami, A., de Mel, G., and Restocchi, V. (2019). Competitive influence maximisation using voting dynamics. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 978–985, New York, NY, USA. ACM.

Chen, W., Lakshmanan, L. V., and Castillo, C. (2013). Information and Influence Propagation in Social Networks. *Synthesis Lectures on Data Management*, 5(4):1–177.

Chen, W., Wang, C., and Wang, Y. (2010a). Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '10*, page 1029, New York, New York, USA. ACM Press.

Chen, W., Wang, Y., and Yang, S. (2009). Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '09*, page 199, New York, New York, USA. ACM Press.

Chen, W., Yuan, Y., and Zhang, L. (2010b). Scalable Influence Maximization in Social Networks under the Linear Threshold Model. In *2010 IEEE International Conference on Data Mining*, pages 88–97. IEEE.

Chen, Y. and Ye, X. (2011). Projection Onto A Simplex. *arXiv*.

Chinellato, D. D., Epstein, I. R., Braha, D., Bar-Yam, Y., and de Aguiar, M. A. M. (2015). Dynamical Response of Networks Under External Perturbations: Exact Results. *Journal of Statistical Physics*, 159(2):221–230.

Chitra, U. and Musco, C. (2020). Analyzing the Impact of Filter Bubbles on Social Network Polarization. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, pages 115–123, New York, NY, USA. ACM.

Choi, D., Chun, S., Oh, H., Han, J., and Kwon, T. (2020). Rumor Propagation is Amplified by Echo Chambers in Social Media. *Scientific Reports*, 10(1):310.

Cioffi-Revilla, C. and Rouleau, M. (2010). Mason rebeland: An agent-based model of politics, environment, and insurgency. *International Studies Review*, 12(1):31–52.

Clifford, P. and Sudbury, A. (1973). A model for spatial conflict. *Biometrika*, 60(3):581–588.

Colaiori, F. and Castellano, C. (2015). Interplay between media and social influence in the collective behavior of opinion dynamics. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 92(4):1–9.

Condie, S. A. and Condie, C. M. (2021). Stochastic events can explain sustained clustering and polarisation of opinions in social networks. *Scientific Reports*, 11(1):1355.

Crokidakis, N. (2012). Effects of mass media on opinion spreading in the Sznajd sociophysics model. *Physica A: Statistical Mechanics and its Applications*, 391(4):1729–1734.

De, A., Bhattacharya, S., and Ganguly, N. (2018). Demarcating Endogenous and Exogenous Opinion Diffusion Process on Social Networks. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*, pages 549–558, New York, New York, USA. ACM Press.

De Domenico, M. and Altmann, E. G. (2020). Unraveling the Origin of Social Bursts in Collective Attention. *Scientific Reports*, 10(1):4629.

De Sio, L. and Weber, T. (2014). Issue Yield: A Model of Party Strategy in Multidimensional Space. *American Political Science Review*, 108(4):870–885.

Deffuant, G., Amblard, F., Weisbuch, G., and Faure, T. (2002). How can extremism prevail? A study based on the relative agreement interaction model. *Journal of artificial societies and social simulation*, 5(4).

Deffuant, G., Neau, D., Amblard, F., and Weisbuch, G. (2000). Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3:87–89.

DeGroot, M. (1974). Reaching a consensus. *Journal of the American Statistical Association*, 69:118–121.

Del Vicario, M., Scala, A., Caldarelli, G., Stanley, H. E., and Quattrociocchi, W. (2017). Modeling confirmation bias and polarization. *Scientific Reports*, 7(December 2016):1–9.

Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G., and Quattrociocchi, W. (2016). Echo Chambers: Emotional Contagion and Group Polarization on Facebook. *Scientific Reports*, 6(1):37825.

Dezső, Z. and Barabási, A.-L. (2002). Halting viruses in scale-free networks. *Physical Review E*, 65(5):055103.

Dhamal, S., Chahed, T., Ben-Ameur, W., and Altman, E. (2018). Optimal multiphase investment strategies for influencing opinions in a social network. *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, 3:1927–1929.

Domingos, P. and Richardson, M. (2001). Mining the network value of customers. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data mining - KDD '01*, pages 57–66, New York, New York, USA. ACM Press.

Dong, Y., Ding, Z., Martínez, L., and Herrera, F. (2017). Managing consensus based on leadership in opinion dynamics. *Information Sciences*, 397-398:187–205.

Downs, A. (1957). An economic theory of political action in a democracy. *Journal of political economy*, 65(2):135–150.

Dragu, T. and Fan, X. (2016). An Agenda-Setting Theory of Electoral Competition. *The Journal of Politics*, 78(4):1170–1183.

Enyedi, Z. and Deegan-Krause, K. (2010). Introduction: The structure of political competition in western europe. *West European Politics*, 33(3):415–418.

Epstein, J. M. and Axtell, R. (1996). *Growing artificial societies: social science from the bottom up*. Brookings Institution Press.

Erdös, P. and Rényi, A. (1959). On random graphs i. *Publicationes Mathematicae Debrecen*, 6:290–297.

Erkol, S., Castellano, C., and Radicchi, F. (2019). Systematic comparison between methods for the detection of influential spreaders in complex networks. *Scientific Reports*, 9(1):1–11.

Eshghi, S., Preciado, V. M., Sarkar, S., Venkatesh, S. S., Zhao, Q., D'Souza, R., and Swami, A. (2020). Spread, Then Target, and Advertise in Waves: Optimal Budget Allocation Across Advertising Channels. *IEEE Transactions on Network Science and Engineering*, 7(2):750–763.

Even-Dar, E. and Shapira, A. (2011). A note on maximizing the spread of influence in social networks. *Information Processing Letters*, 111(4):184–187.

Fekom, M., Vayatis, N., and Kalogeratos, A. (2019). Sequential Dynamic Resource Allocation for Epidemic Control. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 6338–6343. IEEE.

Feld, S. L., Merrill, S., and Grofman, B. (2014). Modeling the effects of changing issue salience in two-party competition. *Public Choice*, 158(3-4):465–482.

Fernández-Gracia, J., Suchecki, K., Ramasco, J. J., San Miguel, M., and Eguíluz, V. M. (2014). Is the Voter Model a Model for Voters? *Physical Review Letters*, 112(15):158701.

Flache, A., Mäs, M., Feliciani, T., Chattoe-Brown, E., Deffuant, G., Huet, S., and Lorenz, J. (2017). Models of Social Influence: Towards the Next Frontiers. *Journal of Artificial Societies and Social Simulation*, 20(4).

Fortunato, S., Latora, V., Pluchino, A., and Rapisarda, A. (2005). Vector opinion dynamics in a bounded confidence consensus model. *International Journal of Modern Physics C*, 16(10):1535–1551.

Fowler, J. H. and Laver, M. (2008). A Tournament of Party Decision Rules. *Journal of Conflict Resolution*, 52(1):68–92.

Fowler, J. H. and Smirnov, O. (2005). Dynamic Parties and Social Turnout: An Agent-Based Model. *American Journal of Sociology*, 110(4):1070–1094.

Friedkin, N. E. and Johnsen, E. C. (1990). Social influence and opinions. *The Journal of Mathematical Sociology*, 15(3-4):193–206.

Fujimoto, T. and Ranade, R. R. (2004). Two characterizations of inverse-positive matrices: the Hawkins-Simon condition and the Le Chatelier-Braun principle. *Electronic Journal of Linear Algebra*, 11(1).

Gaitonde, J., Kleinberg, J., and Tardos, É. (2020a). Adversarial Perturbations of Opinion Dynamics in Networks. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 471–472, New York, NY, USA. ACM.

Gaitonde, J., Kleinberg, J., and Tardos, É. (2020b). Adversarial Perturbations of Opinion Dynamics in Networks. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 471–472, New York, NY, USA. ACM.

Galam, S. (2002). Minority opinion spreading in random geometry. *Eur. Phys. J. B*, 25:403.

Galam, S., Gefen, Y., and Shapir, Y. (1982). Sociophysics: A new approach of sociological collective behaviour. i. mean-behaviour description of a strike. *J. Math. Sociol.*, 9:1–13.

Galam, S. and Jacobs, F. (2007). The role of inflexible minorities in the breaking of democratic opinion dynamics. *Physica A: Statistical Mechanics and its Applications*, 381:366 – 376.

Garcia, D., Mendez, F., Serdült, U., and Schweitzer, F. (2012). Political polarization and popularity in online participatory media. In *Proceedings of the first edition workshop on Politics, elections and data - PLEAD '12*, page 3, New York, New York, USA. ACM Press.

GARCÍA-DÍAZ, C., ZAMBRANA-CRUZ, G., and VAN WITTELOOSTUIJN, A. (2013). POLITICAL SPACES, DIMENSIONALITY DECLINE AND PARTY COMPETITION. *Advances in Complex Systems*, 16(06):1350019.

Ghezelbash, E., Yazdanpanah, M. J., and Asadpour, M. (2019). Polarization in cooperative networks through optimal placement of informed agents. *Physica A: Statistical Mechanics and its Applications*, 536:120936.

Gionis, A., Terzi, E., and Tsaparas, P. (2013). Opinion Maximization in Social Networks. *arXiv e-prints*.

Goldenberg, J., Libai, B., and Muller, E. (2001). Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, 12(3):211–223.

González-Avella, J. C., Eguíluz, V. M., Cosenza, M. G., Klemm, K., Herrera, J. L., and San Miguel, M. (2006). Local versus global interactions in nonequilibrium transitions: A model of social dynamics. *Physical Review E*, 73(4):046119.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. http://www.deeplearningbook.org.

Goyal, M. and Manjunath, D. (2020). Opinion Control Competition in a Social Network. In *2020 International Conference on COMmunication Systems & NETworkS (COMSNETS)*, pages 306–313. IEEE.

Goyal, S., Heidari, H., and Kearns, M. (2019). Competitive contagion in networks. *Games and Economic Behavior*, 113:58–79.

Grabisch, M. and Rusinowska, A. (2020). A Survey on Nonstrategic Models of Opinion Dynamics. *Games*, 11(4):65.

Granovetter, M. (1978). Threshold Models of Collective Behavior. *American Journal of Sociology*, 83(6):1420–1443.

Guare, J. (1990). *Six Degrees of Separation: A Play*. A Vintage original. Vintage Books.

Guimerà, R., Danon, L., Díaz-Guilera, A., Giralt, F., and Arenas, A. (2003). Self-similar community structure in a network of human interactions. *Physical Review E*, 68(6):065103.

Hamill, L. and Gilbert, N. (2009). Social circles: A simple structure for agent-based social network models. *Journal of Artificial Societies and Social Simulation*, 12(2):3.

Han, T. A., Lynch, S., Tran-Thanh, L., and Santos, F. C. (2018). Fostering Cooperation in Structured Populations Through Local and Global Interference Strategies. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, volume 2018-July, pages 289–295, California. International Joint Conferences on Artificial Intelligence Organization.

Hare, C. and Poole, K. T. (2014). The Polarization of Contemporary American Politics. *Polity*, 46(3):411–429.

Hegselmann, R., König, S., Kurz, S., Niemann, C., and Rambau, J. (2015). Optimal opinion control: The campaign problem. *Journal of Artificial Societies and Social Simulation*, 18.

Hegselmann, R. and Krause, U. (2002). Opinion dynamics and bounded confidence: Models, analysis and simulation. *JASSS*, 5(3).

Hegselmann, R. and Krause, U. (2015). Opinion dynamics under the influence of radical groups, charismatic leaders, and other constant signals: A simple unifying model. *Networks & Heterogeneous Media*, 10(3):477–509.

Helbing, D. and Weidlich, W. (1995). Quantitative Sociodynamics - Subject, Methods, Results and Perspectives. *Kolner Zeitschrift Fur Soziologie Und Sozialpsychologie*.

Hoferer, M., Böttcher, L., Herrmann, H. J., and Gersbach, H. (2020). The impact of technologies in political campaigns. *Physica A: Statistical Mechanics and its Applications*, 538:122795.

Hofmann, H., Wickham, H., and Kafadar, K. (2017). Letter-Value Plots: Boxplots for Large Data. *Journal of Computational and Graphical Statistics*, 26(3):469–477.

Holley, R. A. and Liggett, T. M. (1975). Ergodic theorems for weakly interacting infinite systems and the voter model. *The Annals of Probability*, 3(4):643–663.

HOŁYST, J. A., KACPERSKI, K., and SCHWEITZER, F. (2001). SOCIAL IMPACT MODELS OF OPINION DYNAMICS. In *Annual Reviews Of Computational PhysicsIX*, volume 9, pages 253–273. WORLD SCIENTIFIC.

Hu, H. and Zhu, J. J. H. (2017). Social networks, mass media and public opinions. *Journal of Economic Interaction and Coordination*, 12(2):393–411.

Ising, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., and Westwood, S. J. (2019). The origins and consequences of affective polarization in the united states. *Annual Review of Political Science*, 22:129–146.

Jager, W. and Amblard, F. (2005). Uniformity, Bipolarization and Pluriformity Captured as Generic Stylized Behavior with an Agent-Based Simulation Model of Attitude Change. *Computational & Mathematical Organization Theory*, 10(4):295–303.

Jain, M., Jaswani, A., Mehra, A., and Mudgal, L. (2021). EDGly: detection of influential nodes using game theory. *Multimedia Tools and Applications*, (February).

Javarone, M. A. (2014). Network strategies in election campaigns. *Journal of Statistical Mechanics: Theory and Experiment*, 2014(8):P08013.

Jędrzejewski, A. and Sznajd-Weron, K. (2019). Statistical Physics Of Opinion Formation: Is it a SPOOF? *Comptes Rendus Physique*, 20(4):244–261.

Kacperski, K. and Holyst, J. A. (2000). Formation of Opinions under the Influence of Competing Agents — a Mean Field Approach. In *Traffic and Granular Flow '99*, pages 69–80. Springer Berlin Heidelberg, Berlin, Heidelberg.

Kandhway, K. and Kuri, J. (2014). How to run a campaign: Optimal control of SIS and SIR information epidemics. *Applied Mathematics and Computation*, 231:79–92.

Kempe, D., Kleinberg, J., and Tardos, E. (2003). Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, pages 137–146.

Kendall, M. G. (1957). Rank Correlation Methods. *Biometrika*, 44(1/2):298.

Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772):700–721.

Kivela, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., and Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks*, 2(3):203–271.

Kohring, G. A. (1996). Ising Models of Social Impact: the Role of Cumulative Advantage. *Journal de Physique I*, 6(2):301–308.

Kollman, K., Miller, J. H., and Page, S. E. (1992). Adaptive Parties in Spatial Elections. *American Political Science Review*, 86(4):929–937.

Kramer, A. D. I., Guillory, J. E., and Hancock, J. T. (2014). Experimental evidence of massive scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(29):10779–10779.

Krzywinski, M. and Altman, N. (2017). Classification and regression trees. *Nature Methods*, 14(8):757–758.

Kuhlman, C. J., Kumar, V. A., and Ravi, S. (2013). Controlling opinion propagation in online networks. *Computer Networks*, 57(10):2121–2132.

Kurahashi-Nakamura, T., Mäs, M., and Lorenz, J. (2016). Robust Clustering in Generalized Bounded Confidence Models. *Journal of Artificial Societies and Social Simulation*, 19(4).

Kurz, S. (2015). Optimal control of the freezing time in the Hegselmann–Krause dynamics. *Journal of Difference Equations and Applications*, 21(8):633–648.

Kurz, S. and Rambau, J. (2011). On the Hegselmann–Krause conjecture in opinion dynamics. *Journal of Difference Equations and Applications*, 17(6):859–876.

Laguna, M. F., Abramson, G., and Iglesias, J. R. (2013). Compelled to do the right thing. *The European Physical Journal B*, 86(5):202.

Lallouache, M., Chakrabarti, A. S., Chakraborti, A., and Chakrabarti, B. K. (2010). Opinion formation in kinetic exchange models: Spontaneous symmetry-breaking transition. *Physical Review E*, 82:056112.

Langton, C. G. (1995). *Artificial Life: An Overview*. The MIT Press.

Latané, B. (1981). The psychology of social impact. *American Psychologist*, 36(4):343–356.

LAVER, M. (2005). Policy and the Dynamics of Political Competition. *American Political Science Review*, 99(2):263–281.

Laver, M. and Schilperoord, M. (2007). Spatial models of political competition with endogenous political parties. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1485):1711–1721.

Laver, M. and Sergenti, E. (2011). *Party Competition: An Agent-Based Model*. Princeton University Press, Princeton.

Layman, G. C., Carsey, T. M., and Horowitz, J. M. (2006). Party polarization in american politics: Characteristics, causes, and consequences. *Annu. Rev. Polit. Sci.*, 9:83–110.

Lehrer, R. and Schumacher, G. (2018). Governator vs. Hunter and Aggregator: A simulation of party competition with vote-seeking and office-seeking rules. *PLOS ONE*, 13(2):e0191649.

Leskovec, J., Adamic, L. A., and Huberman, B. A. (2007). The dynamics of viral marketing. *ACM Transactions on the Web*, 1(1):5.

Li, G., Motsch, S., and Weber, D. (2020). Bounded confidence dynamics and graph control: Enforcing consensus. *Networks & Heterogeneous Media*, 15(3):489–517.

Li, Y., Chen, W., Wang, Y., and Zhang, Z.-l. (2013). Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships. In *Proceedings of the sixth ACM international conference on Web search and data mining - WSDM '13*, page 657, New York, New York, USA. ACM Press.

Li, Y., Fan, J., Wang, Y., and Tan, K.-L. (2018). Influence Maximization on Social Graphs: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, 30(10):1852–1872.

Lim, S. L. and Bentley, P. J. (2022). Opinion amplification causes extreme polarization in social networks. *Scientific Reports*, 12(1):18131.

Lin, S.-C., Lin, S.-D., and Chen, M.-S. (2015). A learning-based framework to handle multi-round multi-party influence maximization on social networks. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, pages 695–704.

Liu, S., Ying, L., and Shakkottai, S. (2010). Influence maximization in social networks: An ising-model-based approach. In *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 570–576. IEEE.

Liu, Y.-Y. and Barabási, A.-L. (2016). Control principles of complex systems. *Reviews of Modern Physics*, 88(3):035006.

Liu, Y.-Y., Slotine, J.-J., and Barabási, A.-L. (2011). Controllability of complex networks. *Nature*, 473(7346):167–173.

Lorenz, J. (2007). Continuous opinion dynamics under bounded confidence: A survey. *International Journal of Modern Physics C*, 18(12):1819–1838.

Lorenz, J. (2008). Fostering Consensus in Multidimensional Continuous Opinion Dynamics under Bounded Confidence. In *Managing Complexity: Insights, Concepts, Applications*, volume 2008, pages 321–334. Springer Berlin Heidelberg, Berlin, Heidelberg.

Lorenz, J. and Urbig, D. (2007). About the power to enforce and prevent consensus by manipulating communication rules. *Advances in Complex Systems*, 10(02):251–269.

Lu, Q., Korniss, G., and Szymanski, B. K. (2009). The Naming Game in social networks: community formation and consensus engineering. *Journal of Economic Interaction and Coordination*, 4(2):221–235.

Lynn, C. and Lee, D. (2018). Maximizing activity in ising networks via the tap approximation. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*.

Lynn, C. W. and Lee, D. D. (2016). Maximizing influence in an ising network: A mean-field optimal solution. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, pages 2495–2503.

Lynn, C. W. and Lee, D. D. (2017). Statistical mechanics of influence maximization with thermal noise. *EPL (Europhysics Letters)*, 117(6):66001.

Macy, M. W., Ma, M., Tabin, D. R., Gao, J., and Szymanski, B. K. (2021). Polarization and tipping points. *Proceedings of the National Academy of Sciences*, 118(50).

Mai, V. S. and Abed, E. H. (2019). Optimizing Leader Influence in Networks Through Selection of Direct Followers. *IEEE Transactions on Automatic Control*, 64(3):1280–1287.

Majeed, A. and Rauf, I. (2020). Graph theory: A comprehensive survey about graph theory applications in computer science and social networks. *Inventions*, 5(1).

Mandel, A. and Venel, X. (2020). Dynamic competition over social networks. *European Journal of Operational Research*, 280(2):597–608.

Martins, A. C. R. (2008). Continuous opinions and discrete actions in opinion dynamics problems. *International Journal of Modern Physics C*, 19(04):617–624.

Martins, A. C. R. and Galam, S. (2013). Building up of individual inflexibility in opinion dynamics. *Physical Review E*, 87(4):042807.

Marvel, S. A., Hong, H., Papush, A., and Strogatz, S. H. (2012). Encouraging moderation: Clues from a simple model of ideological conflict. *Physical Review Letters*, 109(11).

Mäs, M., Flache, A., and Helbing, D. (2010). Individualization as Driving Force of Clustering Phenomena in Humans. *PLoS Computational Biology*, 6(10):e1000959.

Massa, P., Salvetti, M., and Tomasoni, D. (2009). Bowling alone and trust decline in social network sites. In *Dependable, Autonomic and Secure Computing, 2009. DASC'09. Eighth IEEE International Conference on*, pages 658–663. IEEE.

Mastrandrea, R., Fournet, J., and Barrat, A. (2015). Contact Patterns in a High School: A Comparison between Data Collected Using Wearable Sensors, Contact Diaries and Friendship Surveys. *PLOS ONE*, 10(9):e0136497.

Masucci, A. M. and Silva, A. (2014). Strategic resource allocation for competitive influence in social networks. In *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 951–958. IEEE.

Masuda, N. (2015). Opinion control in complex networks. *New Journal of Physics*, 17:1–11.

Masuda, N., Gilbert, N., and Redner, S. (2010). Heterogeneous voter models. *Phys. Rev. E*, 82:010103.

McFadden, D. (1994). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York.

McFaul, M. and Kass, B. (2019). Understanding Putin's Intentions and Actions in the 2016 U.S. Presidential Election.

Medya, S., Silva, A., and Singh, A. (2019). Influence Minimization Under Budget and Matroid Constraints: Extended Version. *arXiv e-prints*, pages , arXiv:1901.02156.

Meguid, B. M. (2005). Competition Between Unequals: The Role of Mainstream Party Strategy in Niche Party Success. *American Political Science Review*, 99(3):347–359.

Meyer, T. M. and Wagner, M. (2019). It sounds like they are moving: Understanding and modeling emphasis-based policy change. *Political Science Research and Methods*, 7(4):757–774.

Milgram, S. (1967). The small world problem. *Psychology today*, 2:60–67.

Miller, J. H. and Stadler, P. F. (1998). The dynamics of locally adaptive parties under spatial voting. *Journal of Economic Dynamics and Control*, 23(2):171–189.

Mirtabatabaei, A., Jia, P., and Bullo, F. (2014). Eulerian Opinion Dynamics with Bounded Confidence and Exogenous Inputs. *SIAM Journal on Applied Dynamical Systems*, 13(1):425–446.

Mobilia, M. (2003). Does a single zealot affect an infinite group of voters? *Phys. Rev. Lett.*, 91:028701.

Mobilia, M., Petersen, A., and Redner, S. (2007). On the role of zealotry in the voter model. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(08):P08029–P08029.

Molloy, M. and Reed, B. (1995). A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms*, 6(2-3):161–180.

Montes, F., Jaramillo, A. M., Meisel, J. D., Diaz-Guilera, A., Valdivia, J. A., Sarmiento, O. L., and Zarama, R. (2020). Benchmarking seeding strategies for spreading

processes in social networks: an interplay between influencers, topologies and sizes. *Scientific Reports*, 10(1):3666.

Morales, A., Borondo, J., Losada, J., and Benito, R. (2014). Efficiency of human activity on information spreading on Twitter. *Social Networks*, 39:1–11.

Moretti, P., Liu, S. Y., Baronchelli, A., and Pastor-Satorras, R. (2012). Heterogenous mean-field analysis of a generalized voter-like model on networks. *The European Physical Journal B*, 85(3):88.

Moya, I., Chica, M., Sáez-Lozano, J. L., and Cordón, Ó. (2017). An agent-based model for understanding the influence of the 11-M terrorist attacks on the 2004 Spanish elections. *Knowledge-Based Systems*, 123(November 1980):200–216.

Mueller, S. T. and Tan, Y.-Y. S. (2018). Cognitive perspectives on opinion dynamics: the role of knowledge in consensus formation, opinion divergence, and group polarization. *Journal of Computational Social Science*, 1(1):15–48.

Muis, J. (2010). Simulating Political Stability and Change in the Netherlands (1998-2002): an Agent-Based Model of Party Competition with Media Effects Empirically Tested. *Journal of Artificial Societies and Social Simulation*, 13(2).

Muis, J. and Scholte, M. (2013). How to find the 'winning formula'? Conducting simulation experiments to grasp the tactical moves and fortunes of populist radical right parties. *Acta Politica*, 48(1):22–46.

Musco, C., Musco, C., and Tsourakakis, C. E. (2018). Minimizing Polarization and Disagreement in Social Networks. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*, pages 369–378, New York, New York, USA. ACM Press.

Myers, D. G. (1978). Polarizing effects of social comparison. *Journal of Experimental Social Psychology*, 14(6):554–563.

Nayak, A., Hosseinalipour, S., and Dai, H. (2019). Smart Information Spreading for Opinion Maximization in Social Networks. *arXiv e-prints*.

Nesterov, Y. (2004). *Introductory lectures on convex optimization : a basic course*. Kluwer Academic Publishers, Boston.

Newman, M. E. J. (2003). The Structure and Function of Complex Networks. *SIAM Review*, 45(2):167–256.

Nguyen, V. X., Xiao, G., Xu, X.-J., Wu, Q., and Xia, C.-Y. (2020). Dynamics of opinion formation under majority rules on complex social networks. *Scientific Reports*, 10(1):456.

Nishi, R. and Masuda, N. (2013). Collective opinion formation model under Bayesian updating and confirmation bias. *Physical Review E*, 87(6):062123.

Nocedal, J. and Wright, S. J. (1999). *Numerical optimization*. Springer.

Noorazar, H. (2020). Recent advances in opinion propagation dynamics: a 2020 survey. *The European Physical Journal Plus*, 135(6):521.

Noorazar, H., Vixie, K. R., Talebanpour, A., and Hu, Y. (2020). From classical to modern opinion dynamics. *International Journal of Modern Physics C*, 31(07):2050101.

Norman, A. (2016). Why we reason: intention-alignment and the genesis of human rationality. *Biology & Philosophy*, 31(5):685–704.

Nowak, A., Szamrej, J., and Latané, B. (1990). From private attitude to public opinion: A dynamic theory of social impact. *Psychological Review*, 97(3):362–376.

Oestereich, A., Pires, M., Duarte Queirós, S., and Crokidakis, N. (2020). Hysteresis and disorder-induced order in continuous kinetic-like opinion dynamics in complex networks. *Chaos, Solitons & Fractals*, 137:109893.

Osborne, M. (1994). *A course in game theory*. MIT Press, Cambridge, Mass.

Palombi, F., Ferriani, S., and Toti, S. (2017). Influence of periodic external fields in multiagent models with language dynamics. *Physical Review E*, 96(6):062311.

Panovska-Griffiths, J., Swallow, B., Hinch, R., Cohen, J., Rosenfeld, K., Stuart, R. M., Ferretti, L., Di Lauro, F., Wymant, C., Izzo, A., et al. (2022). Statistical and agent-based modelling of the transmissibility of different sars-cov-2 variants in england and impact of different interventions. *Philosophical Transactions of the Royal Society A*, 380(2233):20210315.

Panzarasa, P., Opsahl, T., and Carley, K. M. (2009). Patterns and dynamics of users' behavior and interaction: Network analysis of an online community. *Journal of the American Society for Information Science and Technology*, 60(5):911–932.

Peng, S., Zhou, Y., Cao, L., Yu, S., Niu, J., and Jia, W. (2018). Influence analysis in social networks: A survey. *Journal of Network and Computer Applications*, 106(January):17–32.

Peres, L. R. and Fontanari, J. F. (2011). The media effect in Axelrod's model explained. *EPL (Europhysics Letters)*, 96(3):38004.

Pérez-Llanos, M., Pinasco, J. P., and Saintier, N. (2020). Opinion attractiveness and its effect in opinion formation models. *Physica A: Statistical Mechanics and its Applications*, 559:125017.

Pineda, M. and Buendía, G. (2015). Mass media and heterogeneous bounds of confidence in continuous opinion dynamics. *Physica A: Statistical Mechanics and its Applications*, 420:73–84.

Pineda, M., Toral, R., and Hernández-García, E. (2009). Noisy continuous-opinion dynamics. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(08):P08001.

Porfiri, M. and di Bernardo, M. (2008). Criteria for global pinning-controllability of complex networks. *Automatica*, 44(12):3100–3106.

Proskurnikov, A. V. and Tempo, R. (2017). A tutorial on modeling and analysis of dynamic social networks. Part I. *Annual Reviews in Control*, 43:65–79.

Proskurnikov, A. V. and Tempo, R. (2018). A tutorial on modeling and analysis of dynamic social networks. Part II. *Annual Reviews in Control*, 45:166–190.

Quattrociocchi, W., Caldarelli, G., and Scala, A. (2015). Opinion dynamics on interacting networks: media competition and social influence. *Scientific Reports*, 4(1):4938.

Quattrociocchi, W., Conte, R., and Lodi, E. (2011). Opinions manipulation: Media, power and gossip. *Advances in Complex Systems*, 14(04):567–586.

Rabinowitz, G. and Macdonald, S. E. (1989). A Directional Theory of Issue Voting. *American Political Science Review*, 83(1):93–121.

Rahmattalabi, A., Barman-Adhikari, A., Vayanos, P., Tambe, M., Rice, E., and Baker, R. (2018). Influence maximization for social network based substance abuse prevention. In McIlraith, S. A. and Weinberger, K. Q., editors, *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 8139–8140. AAAI Press.

Ramos, M., Shao, J., Reis, S. D. S., Anteneodo, C., Andrade, J. S., Havlin, S., and Makse, H. A. (2015a). How does public opinion become extreme? *Scientific Reports*, 5(1):10032.

Ramos, M., Shao, J., Reis, S. D. S., Anteneodo, C., Andrade, J. S., Havlin, S., and Makse, H. A. (2015b). How does public opinion become extreme? *Sci. Rep.*, 5:10032.

Rathnayake, C. and Suthers, D. D. (2016). Networked Solidarity: An Exploratory Network Perspective on Twitter Activity Related to #illridewithyou. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pages 2058–2067. IEEE.

Redner, S. (2019). Reality-inspired voter models: A mini-review. *Comptes Rendus Physique*, 20(4):275–292.

Reiljan, A., Kutiyski, Y., and Krouwel, A. (2020). Mapping parties in a multidimensional European political space: A comparative study of the EUvox and euandi party position data sets. *Party Politics*, 26(5):651–663.

Richardson, M. and Domingos, P. (2002). Mining knowledge-sharing sites for viral marketing. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '02*, page 61, New York, New York, USA. ACM Press.

Roman, S., Palmer, E., and Brede, M. (2017). The dynamics of human-environment interactions in the collapse of the classic maya. *Ecological Economics*, 146:312–324. accepted in Ecological Economics.

Romero Moreno, G., Chakraborty, S., and Brede, M. (2021a). Shadowing and shielding: Effective heuristics for continuous influence maximisation in the voting dynamics. *PLOS ONE*.

Romero Moreno, G., Manino, E., Tran-Thanh, L., and Brede, M. (2020a). Zealotry and Influence Maximization in the Voter Model: When to Target Partial Zealots? In *Springer Proceedings in Complexity*, pages 107–118.

Romero Moreno, G., Padilla, J., and Brede, M. (2021b). The effects of party competition on consensus formation. In *3rd International Workshop on Agent-Based Modelling of Human Behaviour (ABMHuB'2021)*.

Romero Moreno, G., Tran-Thanh, L., and Brede, M. (2020b). Continuous influence maximisation for the voter dynamics: is targeting high-degree nodes a good strategy? In *Proc. of the 19th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2020), Auckland, New Zealand, May 9–13, 2020*. IFAAMAS, 3 pages.

Romero Moreno, G., Tran-Thanh, L., and Brede, M. (2020c). Shielding and Shadowing: A Tale of Two Strategies for Opinion Control in the Voting Dynamics. In *Complex Networks and Their Applications VIII. COMPLEX NETWORKS 2019*, volume 881, pages 682–693. Springer, Cham.

Rossi, R. A. and Ahmed, N. K. (2015). The network data repository with interactive graph analytics and visualization. In *AAAI*.

Saito, K., Kimura, M., Ohara, K., and Motoda, H. (2012). Efficient discovery of influential nodes for SIS models in social networks. *Knowledge and Information Systems*, 30(3):613–635.

Schawe, H., Fontaine, S., and Hernández, L. (2021). When network bridges foster consensus. Bounded confidence models in networked societies. *Physical Review Research*, 3(2):023208.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186.

Schweighofer, S., Garcia, D., and Schweitzer, F. (2020a). An agent-based model of multi-dimensional opinion dynamics and opinion alignment. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(9):093139.

Schweighofer, S., Schweitzer, F., and Garcia, D. (2020b). A Weighted Balance Model of Opinion Hyperpolarization. *Journal of Artificial Societies and Social Simulation*, 23(3).

Sen, P. (2011). Phase transitions in a two-parameter model of opinion dynamics with random kinetic exchanges. *Physical Review E*, 83:016108.

Sikder, O., Smith, R. E., Vivo, P., and Livan, G. (2020). A minimalistic model of bias, polarization and misinformation in social networks. *Scientific Reports*, 10(1):1–11.

Sîrbu, A., Loreto, V., Servedio, V. D. P., and Tria, F. (2013a). Cohesion, consensus and extreme information in opinion dynamics. *Advances in Complex Systems*, 16(06):1350035.

Sîrbu, A., Loreto, V., Servedio, V. D. P., and Tria, F. (2013b). Opinion Dynamics with Disagreement and Modulated Information. *Journal of Statistical Physics*, 151(1-2):218–237.

Sîrbu, A., Loreto, V., Servedio, V. D. P., and Tria, F. (2017). *Opinion Dynamics: Models, Extensions and External Effects*, pages 363–401. Springer International Publishing, Cham.

Slawski, M. and Hein, M. (2015). Estimation of positive definite M-matrices and structure learning for attractive Gaussian Markov random fields. *Linear Algebra and its Applications*, 473:145–179.

Smith, J. M. (1976). Evolution and the theory of games: In situations characterized by conflict of interest, the best strategy to adopt depends on what others are doing. *American Scientist*, 64(1):41–45.

Sobkowicz, P. (2015). Extremism without extremists: Deffuant model with emotions. *Frontiers in Physics*, 3(MAR):1–12.

Sobkowicz, P. (2016). Quantitative Agent Based Model of Opinion Dynamics: Polish Elections of 2015. *PLOS ONE*, 11(5):e0155098.

Somer, M. and McCoy, J. (2018). Déjà vu? Polarization and Endangered Democracies in the 21st Century. *American Behavioral Scientist*, 62(1):3–15.

Steels, L. (1995). A Self-Organizing Spatial Vocabulary. *Artificial Life*, 2(3):319–332.

Stewart, A. J., Mosleh, M., Diakonova, M., Arechar, A. A., Rand, D. G., and Plotkin, J. B. (2019). Information gerrymandering and undemocratic decisions. *Nature*, 573(7772):117–121.

Svolik, M. W. (2019). Polarization versus democracy. *Journal of Democracy*, 30(3):20–32.

Tamarit, I., Cuesta, J. A., Dunbar, R. I. M., and Sánchez, A. (2018). Cognitive resource allocation determines the organization of personal networks. *Proceedings of the National Academy of Sciences*, 115(33):8316–8321.

Tavits, M. (2008). Policy Positions, Issue Importance, and Party Competition in New Democracies. *Comparative Political Studies*, 41(1):48–72.

Taylor, M. (1968). Towards a Mathematical Theory of Influence and Attitude Change. *Human Relations*, 21(2):121–139.

Tessone, C. J. and Toral, R. (2009). Diversity-induced resonance in a model for opinion formation. *The European Physical Journal B*, 71(4):549–555.

Tichy, N. M., Tushman, M. L., and Fombrun, C. (1979). Social Network Analysis For Organizations. *Academy of Management Review*, 4(4):507–519.

Toscani, G. (2006). Kinetic models of opinion formation. *Communications in Mathematical Sciences*, 4(3):481–496.

Tran-Thanh, L., Chapman, A., Rogers, A., and Jennings, N. (2021). Knapsack based optimal policies for budget–limited multi–armed bandits. *Proceedings of the AAAI Conference on Artificial Intelligence*, 26(1):1134–1140.

Tsai, J., Nguyen, T. H., and Tambe, M. (2012). Security games for controlling contagion. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence Security*, pages 1464–1470.

Valentini, G., Ferrante, E., Hamann, H., and Dorigo, M. (2016). Collective decision with 100 Kilobots: speed versus accuracy in binary discrimination problems. *Autonomous Agents and Multi-Agent Systems*, 30(3):553–580.

Vassio, L., Fagnani, F., Frasca, P., and Ozdaglar, A. (2014). Message Passing Optimization of Harmonic Influence Centrality. *IEEE Transactions on Control of Network Systems*, 1(1):109–120.

Vaz Martins, T., Pineda, M., and Toral, R. (2010). Mass media and repulsive interactions in continuous-opinion dynamics. *EPL (Europhysics Letters)*, 91(4):48003.

Velásquez-Rojas, F. and Vazquez, F. (2017). Interacting opinion and disease dynamics in multiplex networks: Discontinuous phase transition and nonmonotonic consensus times. *Physical Review E*, 95(5):052315.

Vendeville, A., Guedj, B., and Zhou, S. (2021). Forecasting elections results via the voter model with stubborn nodes. *Applied Network Science*, 6(1):1.

Vendeville, A., Guedj, B., and Zhou, S. (2022). Towards control of opinion diversity by introducing zealots into a polarised social group. In Benito, R. M., Cherifi, C., Cherifi, H., Moro, E., Rocha, L. M., and Sales-Pardo, M., editors, *Complex Networks & Their Applications X*, pages 341–352, Cham. Springer International Publishing.

Vinokur, A. and Burstein, E. (1974). Effects of partially shared persuasive arguments on group-induced shifts: A group-problem-solving approach. *Journal of Personality and Social Psychology*, 29(3):305–315.

Wang, Z., Bauch, C. T., Bhattacharyya, S., D'Onofrio, A., Manfredi, P., Perc, M., Perra, N., Salathé, M., and Zhao, D. (2016). Statistical physics of vaccination. *Physics Reports*, 664:1–113.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442.

Weisbuch, G., Deffuant, G., Amblard, F., and Nadal, J.-P. (2003). Interacting agents and continuous opinions dynamics. In Cowan, R. and Jonard, N., editors, *Heterogenous Agents, Interactions and Economic Performance*, pages 225–242, Berlin, Heidelberg. Springer Berlin Heidelberg.

Wilder, B., Ou, H. C., de la Haye, K., and Tambe, M. (2018a). Optimizing network structure for preventative health. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '18, pages 841–849. International Foundation for Autonomous Agents and Multiagent Systems.

Wilder, B., Suen, S.-C., and Tambe, M. (2018b). Preventing infectious disease in dynamic populations under uncertainty. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*.

Wilder, B. and Vorobeychik, Y. (2018). Controlling elections through social influence. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '18, pages 265–273, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.

Wilford, A. M. (2017). Polarization, number of parties, and voter turnout: Explaining turnout in 26 OECD countries. *Social Science Quarterly*, 98(5):1391–1405.

Wongkaew, S., Caponigro, M., and Borzì, A. (2015). On the control through leadership of the Hegselmann–Krause opinion formation model. *Mathematical Models and Methods in Applied Sciences*, 25(03):565–585.

Xie, X., Li, J., Sheng, Y., Wang, W., and Yang, W. (2021). Competitive influence maximization considering inactive nodes and community homophily. *Knowledge-Based Systems*, 233:107497.

Yadav, A., Wilder, B., Rice, E., Petering, R., Craddock, J., Yoshioka-Maxwell, A., Hemler, M., Onasch-Vera, L., Tambe, M., and Woo, D. (2017). Influence maximization in the field: The arduous journey from emerging to deployed application. In *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*.

Yi, Y. and Patterson, S. (2020). Disagreement and polarization in two-party social networks. *IFAC-PapersOnLine*, 53(2):2568–2575.

Yildiz, E., Ozdaglar, A., Acemoglu, D., Saberi, A., and Scaglione, A. (2013). Binary Opinion Dynamics with Stubborn Agents. *ACM Transactions on Economics and Computation*, 1(4):1–30.

Zhang, H., Vorobeychik, Y., Letchford, J., and Lakkaraju, K. (2016). Data-driven agent-based modeling, with application to rooftop solar adoption. *Autonomous Agents and Multi-Agent Systems*, 30(6):1023–1049.

İkizler, H. (2019). Contagion of network products in small-world networks. *Journal of Economic Interaction and Coordination*, 14(4):789–809.