SCHOOL OF COMPUTER SCIENCE

Swipe, Like, Radicalise?

Incorporating Influencers into Opinion Dynamics: An Agent-Based Model for Understanding Online Polarisation and Extremism

Blaise Sheehan

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Bachelor of Science in the Faculty of Engineering **worth 40CP**.

Friday 25^{th} July, 2025

Dedication and Acknowledgements

To Prof. Seth Bullock, my supervisor, for his much-valued support, guidance, and dedication to this project. His investment in both time and personal expertise, particularly in allowing me, a computer scientist, to deeply explore an aspect of sociology, has been instrumental to the development of this dissertation.

Also to my mum, Prof. Elizabeth Burton, whose research championed the idea of making work not just good, but genuinely impactful. Her approach has been a great source of inspiration throughout this dissertation.

Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Taught Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, this work is my own work. Work done in collaboration with, or with the assistance of others including AI methods, is indicated as such. I have identified all material in this dissertation which is not my own work through appropriate referencing and acknowledgement. Where I have quoted or otherwise incorporated material which is the work of others, I have included the source in the references. Any views expressed in the dissertation, other than referenced material, are those of the author.

Blaise Sheehan, Friday 25th July, 2025

AI Declaration

I declare that any and all AI usage within the project has been recorded and noted within Appendix A or within the main body of the text itself. This includes (but is not limited to) usage of text generation methods incl. LLMs, text summarisation methods, or image generation methods.

I understand that failing to divulge use of AI within my work counts as contract cheating and can result in a zero mark for the dissertation or even requiring me to withdraw from the University.

Blaise Sheehan, Friday 25th July, 2025

Contents

1 Introduction **1**

2 Background **4**
2.1 Social Systems and Social Influence . . . 4
2.2 Networks . . . 4
2.3 Complex and Dynamic Networks . . . 6
2.4 Adaptive Networks . . . 7
2.5 Opinion Dynamics . . . 7
2.6 Opinion Leaders and Influencers . . . 11

3 Base Model **14**
3.1 Structure . . . 14
3.2 Adaptive and Evolutionary Processes . . . 14
3.3 Heterogeneous Parameters . . . 16
3.4 Technical Implementation . . . 17
3.5 Outcome Measures . . . 18
3.6 Data Collation and Analysis . . . 19

4 Extended Model **21**
4.1 Attention Limit . . . 21
4.2 Influencers . . . 22
4.3 Influencer Design . . . 22
4.4 Multiple Influencer Design . . . 24
4.5 Outcome Measures . . . 24
4.6 Data Collation and Analysis . . . 25

5 Base Model - Results **26**
5.1 Outcome Measures . . . 26

6 Extended Model - Results **29**
6.1 Base Model Outcome Measures . . . 29
6.2 Influencer Behaviour . . . 31
6.3 Correlations . . . 34
6.4 Top 50 Neighbour Characteristics . . . 36
6.5 Top 50 Neighbour Opinions . . . 37
6.6 Multiple Influencers . . . 39

7 Discussion **41**
7.1 Influencer Behaviour . . . 41
7.2 Vulnerable Agents . . . 42
7.3 Network Impact . . . 44
7.4 Real-World Consequences and Ideas . . . 45

8 Critical Analysis **47**
8.1 Research Approach . . . 47
8.2 Weaknesses of the Model . . . 47
8.3 Weaknesses of the Influencers . . . 48
8.4 An Alternative View of The Model . . . 49

9 Future Work **51**

10 Conclusion **52**

A Appendix A: Additional Measures **53**

B Appendix B: Artificial Intelligence **55**

List of Figures

2.1 Examples of transition rates $F_{k,m}$ (in terms of the fraction of neighbours m/k holding the opposite opinion) (adapted from Peralta et al., 2022). 9

3.1 Illustration of the set-up and simulation process. 17

4.1 Unweighted network of 4 nodes with varying continuous opinions. Bonds are formed between individuals whose opinions are within 5-points, such as between nodes with opinions of 3 and -1, while no bonds are formed between nodes with opinions of -4 and 3 due to their ideological distance. 21
4.2 Simulation process after the influencers are integrated into the model. The influencers update comes before the network-wide update for both opinions and relationships. 23

5.1 Side-by-side comparison of individual outcome measures between original study and our own simulations of heterogeneous and homogeneous networks ($n = 1000$). 27
5.2 Side-by-side comparison of individual outcome measures between original (top) and new (bottom) simulations of heterogeneous and homogeneous networks ($n = 1000$). 28

6.1 Comparison of outcome measures in non-influencer and influencer networks. The results show similar patterns for both network types, indicating that the presence of influencers does not significantly alter the overall network outcome measures. 30
6.2 Comparison of opinion evolution in networks with and without active influencer behaviour. The results illustrate how influencers significantly alter their opinion dynamics compared to when they act only through regular network interactions. 32
6.3 Comparison of out-strength evolution in networks with and without active influencer behaviour. The results show that when influencers engage in influencer behaviour, their out-strength grows rapidly, whereas in the absence of deliberate behaviour, this increase is weaker, more gradual and may even decline over time. 33
6.4 Evolution of opinion and out-strength over time for Network 13 (light blue) and Network 23 (dark blue). While out-strength increases with opinion extremity, it declines significantly beyond the threshold of -4. 34
6.5 Comparison of the average final out-strength between influencers and regular network nodes in networks with influencers. Influencers have 3.4x greater average out-strength than their peers. 34
6.6 Comparison of correlations between influencer and network opinions and local social norms. 35
6.7 Comparison of conformity, homophily, and attention to novelty parameters for the overall network and the top 50 neighbours of influencers. Highlighting unique social and cognitive characteristics of the most influenced individuals relative to the broader network. 36
6.8 Plot of an influencer's strength across different parameter value pairs, where each dot represents a unique individual indexed by its h_i, a_i value. The x-axis corresponds to the homophily parameter (h) and the y-axis represents the attention to novelty parameter (a). The colour of each dot corresponds to the strength of influence from the influencer, with darker or lighter shades representing varying levels of influence. The influencer's strength varies with these edge characteristics and is greatest for low homophily-high attention to novelty pairs. 37
6.9 Absolute opinion distributions of the top 50 neighbours, comparing scenarios where the influencer is either actively shaping opinions or inactive. 38

6.10 Opinion distances of the top 50 neighbours from the network average in scenarios where the influencer is either actively shaping opinions or inactive. The red plot represents networks with active influencer behaviour, while the blue plot shows networks without it. 38
6.11 Opinion distances of the top 50 neighbours from the influencers opinion, comparing scenarios where the influencer is either actively shaping opinions or inactive. The red plot represents networks with active influencer behaviour, while the blue plot corresponds to networks without it. 39
6.12 Absolute opinion distributions in networks with and without 100 influencers. 40
6.13 Distribution of agent opinions under two conditions: with and without the presence of 100 influencers. Categories are defined relative to the mean and standard deviation of the opinion distribution from the non-influencer network. The inclusion of influencers shifts the opinion landscape, increasing polarisation as evidenced by the rise in extreme (strongly negative and strongly positive) positions. 40

A.1 Plot of final opinion vs. final out-strength for 100 influencers in a network. The out-strength distribution is skewed, reflecting the initial positive bias in opinions. Notably, the final opinion stabilises within the $[6, -6]$ range, rather than the expected $[4, -4]$, due to the reduced neutral population and increased polarisation. 54
A.2 Comparison of assortativity for networks with and without 100 influencers. There is a considerably higher assortativity for multiple influencer networks. 54

Ethics Statement

This project did not require ethical review, as determined by my supervisor, Prof Seth Bullock.

Chapter 1

Introduction

The concept of public opinion emerged during the Enlightenment, when open debate and a marketplace of ideas began to shape Western societies (Brown, 1996). Today, public sentiment influences law, policy, and culture. Even small shifts can drive large socio-political movements, and political scientists routinely rely on public polls to guide decision-making (Judijanto et al., 2024). Understanding how opinions form and evolve is now central to multiple disciplines, with fields like political science, psychology and socio-physics all examining the forces that shape public sentiment.

Opinion formation involves a complex mix of personal traits, social norms, and peer influence. Individuals incorporate social information, whether from opinions or advice, to refine and adjust their own beliefs (Moussaïd et al., 2013). Empirical research in this area includes laboratory experiments with human participants, investigating conformity, authority, social norms, group polarisation, and information exchange in decision-making (Asch, 1956; Haney et al., 1973; Homans, 1958; Sherif, 1936; Stasser and Titus, 1985; Sunstein, 2007). These are complemented by simple theoretical models –such as threshold models, averaging processes, and probabilistic updates- that formalise how beliefs evolve within populations (Degroot, 1974; French and Raven, 1959; Holley and Liggett, 1975). Together, this work shows that opinions rarely form in isolation, but are shaped and reinforced through social interaction.

While much is known about how individuals adjust their views, less is understood about how these micro-level interactions scale up to create societal patterns such as consensus or polarisation (Mason et al., 2007). This gap motivates the study of opinion dynamics. This field uses statistical physics, agent-based modelling, and network theory to explore how opinions change in populations (Hołyst and Kacperski, 2001; Xia et al., 2011). These models serve as computational test-beds for sociological theories, examining how network structures and update rules lead to large-scale outcomes like extremism or collective consensus (Moussaïd et al., 2013). A more thorough introduction to these foundational concepts is provided in Chapter 2.

Early work in this field originated in physics, adapting models such as the Ising model to simulate binary opinion states (Peralta et al., 2022; Sîrbu et al., 2016). This approach laid the groundwork for models of both discrete opinions, where individuals can adopt one of a finite set of opinions (Castellano et al., 2003; Castellano, Muñoz, and Pastor-Satorras, 2009; Granovsky and Madras, 1995), and continuous opinions, where opinions are represented along a spectrum, typically on a real-valued scale, allowing for varied degrees of agreement and disagreement (Hegselmann and Krause, 2002; Nguyen and Shvydkoy, 2023). Researchers employ a range of techniques to model population-level behaviour, including agent-based simulations and abstract mathematical formulations. For the latter, numerical methods, such as mean-field theory, are often used to simplify the models and uncover key relationships between parameters and outcomes (Peralta et al., 2022).

The mechanisms that govern opinion change vary widely - from simple averaging of neighbouring opinions (Degroot, 1974; Hegselmann and Krause, 2002) to imitation processes (Vazquez et al., 2019). More recent models have expanded to incorporate contemporary phenomena, such as algorithmic content curation (Perra and Rocha, 2018), the influence of social media personalities (Coculescu et al., 2023), media strategies (Brooks and Porter, 2020), and optimal opinion placement (Hegselmann et al., 2015). These advancements reflect the increasing complexity of opinion dynamics in the digital age.

Despite their sophistication, many models lack empirical validation and rely on untested assumptions about how people form opinions, limiting their real-world applicability (Carpentras et al., 2022; Castellano, Muñoz, and Pastor-Satorras, 2009; J. Dong et al., 2024; Moussaïd et al., 2013; Peralta et al., 2022; Sobkowicz, 2009)). A key issue is that empirical studies often emphasise accuracy relative to how they implement hypotheses within their models, whereas theoretical research typically focuses on abstract, often mathematical, representations of opinion change. This divide results in a field that is more focused on theoretical modelling than on providing practical insights for social science (Chattoe-Brown, 2022; Sobkowicz, 2009)). In fact, empirical studies are significantly less cited than theoretical ones (Chattoe-Brown, 2022). Additionally, researchers often rely on a few simplified hypotheses for justification, resulting in overextended papers that try to address too many points simultaneously (Carpentras, 2023; Sobkowicz, 2009)). This not only reduces readability and targeted impact but also undermines interdisciplinary efforts in a field that is so clearly interdisciplinary (Carpentras, 2023; Sobkowicz, 2009)).

Nonetheless, the role of technology in social science is evolving. The rise of computational social science and large-scale social media data offers new ways to analyse opinion formation in real-world settings (Lazer et al., 2009; Liu, 2012; Moussaïd et al., 2013; Peralta et al., 2022). Experimental efforts have also begun testing the simplified hypotheses used in these models. One such hypothesis involves the number of dissenters required to sway a group, a key dynamic in the voter model discussed in Chapter 2. These developments raise hope for greater empirical grounding in future work (Carpentras, 2023; Sobkowicz, 2009)).

A key aspect of the digital transformation is the rise of social media influencers (SMIs). While definitions vary across the literature, SMIs are generally understood as individuals with a substantial online presence or following (Duffy, 2020; Enke and Borchers, 2019; Goodwin et al., 2023a). They cultivate large audiences by projecting a brand of perceived authenticity and relatability (Goodwin et al., 2023a; Prawira et al., 2024). This constructed sense of trust enables influencers to monetise their platforms, often through product endorsements and sponsorships, under the guise of personal recommendation (Duffy, 2020). The implicit logic is compelling: if the influencer genuinely loves a product, their followers are likely to as well.

The influence of SMIs extends beyond consumer behaviour into the political sphere. Their established audiences provide political actors with a ready-made, trust-based communication channel, offering a shortcut to engagement that would otherwise require significant time and resources to build (Borchers, 2025; Fischer et al., 2022; Goodwin et al., 2023a; Wong, 2020). For example, in the 2020 U.S. Election Donald Trump, Kamala Harris and Bernie Sanders all collaborated with influencers to boost their campaign reach (Glazer and Wells, 2019). In this sense, influencers operate as brokers of social capital, converting perceived authenticity into economic and, sometimes, political value.

Some influencers specialise in political content, while others in lifestyle domains unintentionally influence political thought through their posts on issues such as health, housing, and social justice (Goodwin et al., 2023a). Their messaging, often more relatable and less overtly partisan than traditional political content, can be especially effective in shaping political attitudes across demographic boundaries (Marriott, 2024). For instance, appearances of political figures on non-political platforms, like Donald Trump on Joe Rogan's podcast, allow for a more humanised, authentic connection with voters, which has proven highly persuasive in comparison to conventional campaign strategies (Marriott, 2024).

Research shows that influencers help boost political efficacy among their followers (the belief that one can influence political processes). They also drive political mobilisation, especially among younger generations (Harff and Schmuck, 2023; Seeger and Muth, 2023). Parasocial relationships—where followers feel a one-sided, personal connection with influencers—further increase the persuasive power of their messages. Influencers then often exploit these emotional bonds to push consumerist or sociopolitical agendas. (Cheng et al., 2023; Stewart et al., 2024).

As social media continues to expand, the influence of these digital figures is only expected to intensify. A significant portion of younger individuals now cite SMIs as their primary source of political information, emphasising the importance of this emerging form of digital political influence (Stewart et al., 2024). This shift underscores the need for new models that better capture the role of influencers in shaping opinion dynamics. Traditional models, which often rely on homogeneous agent structures, fail to account for the significance of influencers in driving attitude changes online.

The surge in online political polarisation, fuelled by the proliferation of extremist content and algorithmic amplification, is connected with the rise of influencers and social media dynamics. Algorithms prioritise content that aligns with users' existing beliefs, creating feedback loops that deepen ideological divides and reinforce selective exposure (Gentzkow and Shapiro, 2018; Huszár et al., 2022; Muñoz et al., 2024; Peralta et al., 2022; Prior, 2013). This selective exposure contributes to the formation of echo chambers, where individuals are predominantly exposed to like-minded perspectives, thereby intensifying political polarisation (Bail, Argyle, Brown, Bumpus, et al., 2018; Baumann, Lorenz-Spreen, et al., 2020; Levy, 2021; Vicario et al., 2016).

Moreover, the influence of SMIs has disrupted traditional notions of opinion leadership. Unlike conventional media outlets that often maintain editorial standards, influencers frequently lack the same level of credibility, with many failing to verify the accuracy of the information they disseminate. A UNESCO study found that 62% of influencers do not fact-check the content they share, relying instead on metrics such as likes and views to gauge the appeal of their posts (Ha, 2024). This lack of rigorous verification contributes to the rapid spread of misinformation, which is a significant driver of political polarisation (Chen et al., 2021; Muñoz et al., 2024).

While studies have found that influencer extremism leads to greater follower polarisation, the evidence is mixed when ascertaining the exact impact of influencers on driving global political polarisation. The dynamics are complex, with some studies reiterating a strong correlation between extremist content and polarisation, while others suggest that the exact influence of influencers may not be as substantial as initially assumed (Gentzkow et al., 2018; Kosse and Carstens, 2023).

This dissertation seeks to address a critical gap in existing opinion dynamics models by incorporating influencers into a formal framework, thereby offering both computational and sociological insights into how influencers shape collective opinion. By empirically linking theoretical insights with known behaviours, the dissertation offers a more grounded understanding of the dynamics of influencers in the digital age. Through the use of agent-based modelling, we can explore micro-level behaviours often overlooked by more abstract numerical approaches. Similarly, by focusing specifically on influencers, we avoid the risk of overextending our research across too many factors. This approach helps maintain a strong connection to social science and ensures that technical concepts are properly justified, unlike in other papers like Carpentras (2023). We therefore move beyond abstraction, providing insights that are directly relevant to the impact of influencers on public opinion today.

Key Objectives

1. Situate and justify our research within the existing body of opinion dynamics literature, focusing on the role of influencers in shaping opinion dynamics. We will clarify the current state of the field while providing a strong rationale for the relevance and importance of our project's scope.

2. Motivate the choice of our model, explaining why it serves as an appropriate foundation for our work. Next, replicate the findings of the original heterogeneous model by Bullock and Sayama (2023). Then, extend this model by incorporating empirically validated social phenomena, processes, and ideas to further ground the model in real-world dynamics.

3. Define a computationally realistic concept of an influencer that (1) reflects real-world influence dynamics, (2) remains lightweight enough to preserve computational efficiency, and (3) allows agents strategic freedom in how they pursue influence maximisation within the population.

4. Thoroughly analyse and interpret key findings within a social context, aiming to (1) establish connections to the real world where relevant, and (2) use mathematical explanations and analyses to differentiate between genuine dynamics and artifacts of computational constraints.

5. Provide a thorough critical evaluation of the model, identifying (1) weaknesses and limitations, and (2) offering constructive suggestions for improvement and potential directions for future research.

Chapter 2

Background

2.1 Social Systems and Social Influence

Social systems are composed of individuals whose behaviours, beliefs, and decisions are shaped through ongoing interactions with others. These systems, ranging from families to country-level populations, exhibit collective behavioural patterns that cannot be fully explained by looking at individuals in isolation. Understanding such processes sheds light on how widespread behaviours form and evolve, from common phenomena like social norms and cultural trends to more volatile events such as social prejudice, riots, and collective fear during epidemics (Moussaïd et al., 2013). This line of inquiry falls within social psychology, the scientific study of how people's thoughts, feelings, and actions are influenced by the real or perceived presence of others (Allport, 1954).

A core driver of collective behaviour is social influence, including processes like conformity, peer pressure and obedience. Earlier studies of these processes included social norms (Sherif, 1936), obedience (Milgram, 1964), social conformity (Asch, 1956; Turner, 1985, 1991), herd behaviour (Keynes, 1936), minority influence (Moscovici and Nève, 1971) (in which the impact of the opinions of a minority is considered), workplace influence (French and Raven, 1959), social prejudices (Kelley and Volkart, 1952) and the role of social institutions (Bogardus, 1924). From these studies, various theories of social influence have evolved, mostly differentiated by their ideas on how and why humans are so deeply subject to the influence of their social environment (Rashotte, 2007).

Building on these foundations, research has increasingly focused on the role of social influence in shaping public opinion. This work spans empirical and theoretical studies. On the empirical side, lab experiments and surveys have studied how individuals adjust opinions in response to peer interaction and social pressure (Asch, 1956; Haney et al., 1973; Homans, 1958). On the theoretical side, formal models have provided mathematical ways of understanding how opinions are shared within networks of humans (Degroot, 1974; French and Raven, 1959; Moussaïd et al., 2013). Collectively, these approaches reiterate that social influence drives large-scale opinion dynamics, producing outcomes like group consensus and political polarisation.

2.2 Networks

Network science traces its origins to Leonard Euler's 1736 work on the Koenigsberg Bridges, where he represented the city's bridges and landmasses as a network of nodes and edges (Euler, 1736). The application of this structure to a wide range of systems has facilitated the analysis of various social, biological and physical networks within a mathematical framework. For instance, one could model European cities as nodes and direct flight routes as edges, or represent organs as nodes with blood vessels forming the connecting edges. The rise of powerful computing resources has since enabled temporal simulations and the complex analysis of such networks-tasks that are typically computationally intensive. Today, network science underpins a wide array of fields, from mapping terrorist networks (Sureda et al., 2017) and developing new approaches to mental health treatment (Hofmann and Curtiss, 2018), to analysing global air travel patterns (Huynh et al., 2024) and simulating protein interactions at the molecular level (Safari-Alighiarloo et al., 2014).

2.2.1 Formal Definition

A network is defined as a set of nodes V with edges E between them.

$$G = (V, E)$$

such that:

- The edges E can be **undirected** or **directed**. In an undirected graph, an edge between two nodes i and j is represented as (i, j), meaning the nodes are associated with each other but not directionally. In a directed graph, an edge is denoted as $(i \to j)$, where the connection is from node i to node j.
- The edges can also be **weighted** or **unweighted**. In a weighted graph, each edge $e \in E$ is associated with a numerical value w, representing the strength of an edge between two nodes. That is,

  $$w : E \to \mathbb{R}$$

  For **directed, weighted graphs**, we use the notation w_{ij} to represent the weight of an edge from node i to node j, where:

  $$w_{ij} \neq w_{ji}$$

Thus, a weighted, directed graph is formally represented as:

$$G = (V, E, W)$$

where W is the set of weights for each directed edge w_{ij}.

2.2.2 Neighbours

A **neighbour** of a node i is any other node j that is directly connected to i by an edge. Formally, for a graph $G = (V, E)$, a node $i \in V$ is a neighbour of node $j \in V$ if there exists an edge $(i, j) \in E$.

2.2.3 Communities

A **community** in a graph $G = (V, E)$ is a subset of nodes $C \subseteq V$ such that the density of edges within C is significantly higher than the density of edges between C and the rest of the graph $V \setminus C$ (Fortunato, 2009).

Letting w_{ij} denote the weight of the edge from node j to node i. Then C is a community if:

$$\sum_{i,j \in C} w_{ij} \gg \sum_{\substack{i \in C \\ j \in V \setminus C}} w_{ij}$$

Here, the notation $\gg$ (read as much greater than) signifies that the total edge weight within the community C is substantially larger than the total edge weight connecting C to the rest of the network. This reinforces the idea that communities have greater internal cohesion than external connectivity.

These communities (sometimes called clusters) often represent functional or meaningful subgroups within a network, such as social circles in a social network or clusters of related web pages on the internet.

2.2.4 Modularity

Let $G = (V, E)$ be a graph with $n = |V|$ nodes, and let A be the adjacency matrix where A_{ij} is the weight of the edge between nodes i and j. Let $C = \{C_1, C_2, \ldots, C_k\}$ be a partition of V into k communities. The modularity Q of this partition is:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(x_i, x_j)$$

Where: $m = \frac{1}{2} \sum_{i,j} A_{ij}$ is the total weight of the edges. $k_i = \sum_j A_{ij}$ is the degree of node i. $\delta(x_i, x_j)$ is the Kronecker delta function, which is one if i and j belong to the same community, and zero otherwise. $\frac{k_i k_j}{2m}$ is the expected edge weight between nodes i and j in a random graph with the same degree distribution.

The modularity equation works by, for every pair of nodes, comparing the actual edge weight with the expected edge weight for a random graph. If the nodes belong to the same community, the difference between the actual and expected edge weights is multiplied by one (i.e., $\delta(x_i, x_j) = 1$). Modularity increases when the difference between the actual and expected is large, and the nodes are in the same community. If the difference is small, the contribution to modularity is minimal. If the nodes are in different communities, they are ignored, meaning no contribution to modularity.

Higher modularity is thus the result of nodes belonging to the same community, whose edge weight between each-other is much greater, and thus more densely connected, than in a randomly distributed graph.

2.2.5 Community Detection Algorithms

Community detection algorithms are techniques used to identify such communities (Fortunato, 2009). One popular method for community detection is the Louvain algorithm, which is an efficient approach based on modularity optimisation.

Louvain Method

The algorithm works in two phases:

1. Each node is assigned its own community. Then, for each node, it checks the change in modularity if the node were to move to every other neighbour's community. The node is moved into whichever community, greedily, results in the highest modularity gain.

2. After all nodes are processed, communities become the nodes and Step (1) is repeated. Such that sub-communities are moved into other sub-communities. This continues until no further improvements to modularity can be made.

This method has a time complexity of $O(m \log n)$, making it one of the fastest community detection algorithms for large networks with millions of nodes and edges (Blondel et al., 2008).

2.2.6 Out-Strength

We define the out-strength of node j as:

$$\textbf{Out-strength of node } j = \sum_{i \in V} w_{ij}$$

where w_{ij} represents the edge from node j to node i, and V is the set of all nodes in the network. Intuitively, this is total weight of node j's out-going connections.

2.2.7 In-Strength

We define the in-strength of node j as:

$$\textbf{In-strength of node } j = \sum_{i \in V} w_{ji}$$

where w_{ji} represents the edge from node i to node j, and V is the set of all nodes in the network. Intuitively, this is the total weight of all the incoming connections to node j.

2.3 Complex and Dynamic Networks

Complex networks refer to networks that exhibit non-random, organised structures. These networks better represent real-world systems where relationships between nodes follow complex design principles rather than random connections. For example, in a Facebook friendship network, we expect to see tightly-knit communities, as individuals often befriend others within their social circles or geographic regions. This pattern of clustering is an example of a non-random process influencing the structure of a network.

In our airport example, regional hubs form as nearby cities are more likely to be directly connected due to geography and travel demand, creating clusters driven by real-world patterns rather than randomness.

Dynamic networks expand on complex networks by introducing the element of time, allowing the structure and/or state of the network to evolve. They are studied in two main contexts:

1. **Dynamics of the network:** This refers to how the network's topology (edges) changes over time. For example, new friendships could form, or existing relationships might break, altering the connections between individuals in a social network.

2. **Dynamics on the network:** This refers to how the state of the nodes within the network evolves over time. For example, individuals becoming infected in an epidemic network or an airports daily passenger volume increasing overtime.

Earlier studies often focused on a single aspect. Either the evolution of network connections (Barabási and Albert, 1999; D. J. Watts and Strogatz, 1998) or changes in node states within a static network (Bullock and Sayama, 2023; Gross and Blasius, 2007; Sayama et al., 2013; Sood and Redner, 2005). More recent approaches combine both, capturing the dynamic interplay between connections and states.

2.4 Adaptive Networks

The study of **adaptive networks** is thus the study of networks whose topology is interlinked with the node states. Changes in topology affect node states and changes in the states cause changes in the topology, creating a feedback loop. This dynamic has become a rich research area (Dorogovtsev et al., 2008; Gross and Blasius, 2007; Sayama et al., 2013) with applications including ecology, social systems, and engineering (Antoniades and Dovrolis, 2015; Baumann, Lorenz-Spreen, et al., 2020; Berner et al., 2021; Proulx et al., 2005).

For example, consider an epidemic network model where the nodes represent individuals, each with a state such as infected or healthy, and the edges represent their interactions or contacts. If an individual becomes infected (state changes to infected), their connection to a healthy individual may be removed, such as through social distancing measures that limit contact. This reflects a change in the network's structure or topology. Likewise, if a healthy individual comes into contact with an infected person, such as by meeting a new friend, a new edge is formed in the network. This new connection increases the chance of transmission, and the healthy individual may become infected as a result, changing their state from healthy to infected. Clearly, there is an interplay between changes in node states and changes in the network's structure, with each influencing the other.

2.5 Opinion Dynamics

Adaptive networks are ideal for modelling opinion dynamics, as they capture the feedback loop between social influence and individual opinions, where changes in opinions shape individual connections and vice versa.

Formally, **opinion dynamics** uses quantitative frameworks and models to replicate how opinions spread and evolve within populations. These models offer insights for various fields including marketing (Martins et al., 2009), finance (Kaizoji, 2006; Kaizoji et al., 2002), politics (Bernardes et al., 2001; Y. Dong et al., 2020) and public health (Xia et al., 2011). The field has also explored the roles of specific agents in shaping public opinion, such as influencers (Brooks and Porter, 2020; Chazelle, 2012; Helfmann et al., 2023), mass media (Brooks and Porter, 2020), and disruptively stubborn individuals like zealots (Mobilia et al., 2007; Verma et al., 2014).

Rooted in statistical physics, the field of opinion dynamics, often referred to as socio-physics, has significant gaps between its methods and those of traditional social science (Sobkowicz, 2009)). Despite opinion formation being clearly rooted in social sciences, most studies rely on complex numerical analyses with limited empirical validation, making the field difficult for social scientists to engage with effectively. This disconnect hampers meaningful interdisciplinary collaboration, limiting the integration of valuable social science insights into model development and vice versa (Carpentras et al., 2022; J. Dong et al., 2024; Moussaïd et al., 2013).

Existing models can be divided into two types:

1. **Discrete Opinion Models**: These models consider opinions as binary choices, such as agreement or disagreement on an issue or acceptance of an idea (Cao et al., 2024; Castellano et al., 2003;

Castellano, Muñoz, and Pastor-Satorras, 2009; Granovsky and Madras, 1995; Lian and Dong, 2022; Sznajd-Weron and Sznajd, 2000).

2. **Continuous Opinion Models**:These models consider opinions as points on a spectrum, allowing them to represent an individual's stance on an issue(Bullock and Sayama, 2023; den Heijer et al., 2024; Hegselmann and Krause, 2002; Nguyen and Shvydkoy, 2023; Sayama, 2020; Weisbuch et al., 2002).

Most models employ an update mechanism to determine how an individual's opinion evolves over time. These mechanisms range from simple opinion copying to iterative methods that use weighted relationships to update opinions.

Continuous opinion models can be further classified into **linear** and **non-linear** models based on the update mechanism employed.

1. **Linear models**: The update mechanism directly relates an individual's opinion to external interactions, typically assuming that opinion changes are proportional to influences from others.

2. **Non-linear models**: The update mechanism also includes thresholds, parameters or other non-linear dependencies and relations, making opinion evolution more complex (Hegselmann and Krause, 2002). This adds complexity to the opinion formation processes.

Linearity effectively captures the different theories on social interaction and opinion formation.

2.5.1 A General Framework

We begin by outlining a general framework that underpins most opinion dynamics models based on network and agent-based approaches. This excludes models formulated solely through differential equations without simulation components.

Consider a network $G = (V, E)$, where the set of nodes V represents agents (i.e. individuals), each holding an opinion value x_i, which may be either discrete or continuous. The edges E represent interactions or influence between agents and can be either weighted or unweighted—weights indicating the strength of influence, while unweighted edges simply indicate the presence of a connection or interaction. Edges may also be directed or undirected. A directed edge reflects asymmetric influence, meaning one individual may influence another without being equally influenced in return. In contrast, undirected edges assume mutual influence, where both individuals equally affect each other.

There is also an update mechanism governing how an individual's opinion changes with each time step t. Formally, the opinion of an individual i at time t, denoted as $x_i[t]$, is updated based on its neighbours.

$$x_i[t+1] = f(x_i[t], \{x_j[t] \mid j \in N(i)\})$$

where $N(i)$ represents the neighbours of node i, and f is a function that describes how an individual's opinion is influenced by others. The specific form of f varies depending on the model being used but is mostly rooted in differing sociological theories.

2.5.2 Discrete Models

Discrete models use binary opinion states, that is the variable x_i representing an individual's opinion is such:

$$x_i \in \{-1, 1\}$$

where $x_i = -1$ corresponds to support for Party A or rejection of some idea and $x_i = 1$ corresponds to support for Party B or acceptance of some idea.

The use of binary opinions simplifies the mathematics behind these models (Peralta et al., 2022), enabling less complex analyses of outcomes.

We define transition rates, which represent the probability per unit time that an individual will change their opinion between {-1, 1}. This probability is denoted as $F_{k,m}$, where m represents the number of opposing neighbours, and k is the total number of neighbours (Gleeson, 2011). Thus, each node updates its opinion based on probabilities dependent on its neighbours opinions.

The shape of $F_{k,m}$ in Figure 2.1 represents different hypotheses on social influence. The dashed purple line demonstrates the widely used Voter Model (Castellano, Muñoz, and Pastor-Satorras, 2009). Here, the probability of a node's opinion switching is equal to the number of individuals in its neighbourhood holding that opinion. Several extensions of this voter model exist where transition rates are non-linear (Castellano, Muñoz, and Pastor-Satorras, 2009) i.e. the red and blue conformity lines in Figure 2.1. Interestingly, empirical work is being done to validate the specific shape of such transition rates in human-based laboratory experiments (Chacoma and Zanette, 2015; Vande Kerckhove et al., 2016).

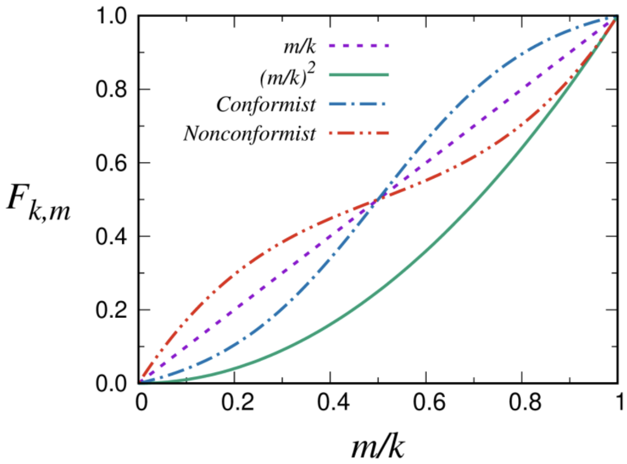


Figure 2.1: Examples of transition rates $F_{k,m}$ (in terms of the fraction of neighbours m/k holding the opposite opinion) (adapted from Peralta et al., 2022).

Work has also been done in multi-opinion systems, that is:

$$x_i \in S = \{s_1, s_2, \ldots, s_k\}$$

for k different opinions (Herrerías-Azcué and Galla, 2019; Vazquez et al., 2019)(i.e. k different political parties, or k discrete stances on an opinion). Interestingly, models with binary and multiple discrete states behave very similarly (Peralta et al., 2022). When scaling is applied, the two models end up with equivalent statistical results, suggesting that the binary model is sufficient in studies on discrete opinions.

The Voter Model

The voting problem originated in probability theory in the 1970s as a study of social competition (Clifford and Sudbury, 1973) and was later extended to explore the dynamics of public voting behaviour (Holley and Liggett, 1975). It eventually evolved into a formal mathematical model, representing a population first on a lattice structure and later as a network (Castellano et al., 2003).

At each time step, the model randomly selects a neighbour, and the agent adopts that neighbour's opinion (Clifford and Sudbury, 1973). This reflects the transition rate $F_{k,m} = \frac{m}{k}$, as the likelihood of switching depends on the proportion of neighbours holding the opposite opinion. Over time, this process drives the network towards consensus, with the majority opinion eventually dominating.

Sznajd extended the model by proposing that an individual is more likely to be influenced by a pair of neighbours than by a single one (Stauffer, 2002; Sznajd-Weron and Sznajd, 2000). Although this extension has faced criticism for its lack of empirical grounding (Behera and Schweitzer, 2003), the Sznajd model remains one of the most studied in discrete opinion dynamics (Xia et al., 2011).

Despite the Voter Model's simplistic portrayal of human behaviour, where copying dynamics, though partially empirically supported, are often overused, its direct applicability to real-world scenarios remains limited (Stafford and Ekland-Olson, 1982). However, empirical extensions of the model have shown greater realism, successfully replicating phenomena such as voting patterns in the 2020 U.S. election (Bouchaud et al., 2014).

2.5.3 Continuous Models

Continuous opinion models represent individuals' opinions as real-valued variables, allowing for a gradual spectrum of beliefs compared to discrete models. A heavily studied subset of these are bounded-confidence models, originally developed by Deffuant (Deffuant et al., 2000) (D-W) and Hegselmann-Krause (Hegselmann and Krause, 2002) (H-K).

Both models use bounded confidence: agents influence each other only if their opinion difference is below a threshold ϵ. This idea appears in various models under different terms, such as receptiveness or, as we later define, attention limits (Brooks and Porter, 2020).

When the threshold ϵ is large (indicating a high level of receptiveness for different opinions), the system tends towards **consensus**, a global state where all individuals have a similar opinion value. For low thresholds (indicating a low level of tolerance for different opinions), **polarisation** occurs, a state where two distinct opinion groups emerge. If ϵ is very small, fragmentation happens, leading to a state where three or more distinct opinion groups form.

Both the Hegselmann-Krause (H-K) and Deffuant-Weisbuch (D-W) models share a common update mechanism: agents adjust their opinions based on the opinions of those within their confidence bounds. However, they differ in how this adjustment is implemented. In the H-K model, agents update their opinions simultaneously based on all neighbours within the threshold, while in the D-W model, updates occur through random pairwise interactions, making the process more stochastic in nature.

Deffuant-Weisbuch Model (D-W)

The Deffuant-Weisbuch model (Deffuant et al., 2000; Weisbuch et al., 2002) focusses on **pairwise** interactions, where at each time step t, a randomly chosen agent i interacts with another randomly chosen neighbour j if their opinions satisfy

$$|x_i[t] - x_j[t]| < \epsilon.$$

If this condition holds, agent i updates its opinion according to the rule

$$x_i[t+1] = x_i[t] + c(x_j[t] - x_i[t]),$$

where $c \in (0, 1]$ is a conformity parameter that controls how strongly an agent moves towards its neighbour's opinion, that is, how conformist node i is. The agent i is therefore unconcerned with the network majority.

This D-W model has been used to study extremist opinions in society (Marconi and Cecconi, 2020) and the effects of persuasion (Weisbuch et al., 2004). The pairwise approach makes Deffuant's model better suited to opinion dynamics in large populations because it allows for localised interactions between individuals, reducing the computational complexity compared to models that require global interactions (Castellano, Fortunato, and Loreto, 2009).

DeGroot's Model

The DeGroot averaging mechanism dictates how individuals change their opinions within a network with weighted, directed edges (Degroot, 1974). This is the mechanism used in the H-K model. The opinion $x_i(t) \in [0, 1]$ evolves according to:

$$x_i(t+1) = \sum_{j=1}^{N} w_{ij} x_j(t), \tag{2.1}$$

where w_{ij} are normalised weights such that:

$$\sum_{j=1}^{N} w_{ij} = 1. \tag{2.2}$$

In simple terms, each individual's opinion at the next time step is a weighted average of the opinions of others in the network, where the weights reflect the strength of influence or connection between individuals.

This approach reduces a probability problem (recall transition rates) into a simpler, algebraic problem and the dynamics of opinion spread then rely on the topology of the network itself rather than the shape of the opinion change function. It also reinforces the idea that social interaction is crucial in individual opinion formation.

Hegselmann-Krause Model (H-K)

In contrast to the D-W model, the H-K model (Hegselmann and Krause, 2002) considers **all neighbours** within the confidence threshold. At each time step, an agent i updates its opinion with Eq. 2.1, averaging only over neighbours that satisfy the boundary.

Specifically, an agent i considers a set of neighbours whose opinions are within a confidence threshold ϵ. This set is defined as:

$$N_i^\epsilon = \{j \mid |x_i(t) - x_j(t)| < \epsilon\}. \tag{2.3}$$

The opinion update then follows the DeGroot mechanism but is restricted to this subset of neighbours.

The H-K model has been extensively studied, including research on strategic oracles (Rodrigo, 2025) (artificially intelligent agents that, after digesting relevant information, answer questions much like Large Language Models such as OpenAI's ChatGPT) and heterophilic relationships (Motsch and Tadmor, 2014). Unlike the D-W model, the H-K model is best suited to formal meetings or gatherings because it assumes that all individuals interact simultaneously and deliberately. This allows us to model environments where individuals engage in collaborative, direct, often centralised, exchanges of views, unlike the more decentralised interactions in the D-W model (Castellano, Fortunato, and Loreto, 2009).

2.6 Opinion Leaders and Influencers

Long before the digital era, researchers recognised that public opinion is the result of complex social norms and processes. Katz and Lazarsfeld's seminal two-step flow model demonstrated that top-level information typically flows first from mass media to a small group of opinion leaders, who then shape how that information spreads to the public (Katz and Lazarsfeld, 1955). These individuals act as crucial intermediaries by filtering, re-framing, and selectively amplifying messages based on their more personal influence.

The concept of the opinion leader traditionally sees the individual as embedded within a structured social group. These groups are defined by strong personal or physical connections, such as those between a parent and child, an employee and employer or a professor and their students. Each of these examples reflects the assumption that the opinion leader holds a position of greater credibility or trust within their social group, which enables them to influence how information is received and interpreted.

At the time of the Two-Step Flow Theory's development, mass communication was largely linear. For example, a news outlet might broadcast a story, which is then interpreted and passed on by a mother to her family. In this, the message is not merely relayed but filtered through her perspective, subtly shaping how her family receives and understands the information.

In the decades since, the media landscape has undergone a profound transformation. The rise of social media has dramatically expanded the channels through which individuals access information. Increasingly, people cite social media as their primary source of news and current affairs (Perrin, 2015; Williams, 2017). This decentralisation of information dissemination has thus given rise to a new class of opinion leaders: social media influencers. Whilst the definition of influence is not clearly defined, studies typically assume it is accounts who amass large followings online. These individuals build these audiences through a combination of curated content, perceived authenticity, and direct engagement with followers (Borchers, 2025; Hugh et al., 2022).

In this way, the Two-Step Flow Theory remains relevant, albeit in a transformed context. Where once the opinion leader might have been a community elder or family member, today they may be a lifestyle vlogger or Instagram personality. Yet the core dynamic persists: information continues to pass through a trusted intermediary, someone perceived as more credible or authoritative, before reaching the wider group. As such, the theory can be reframed for the digital age, with influencers stepping into the role of digital opinion leaders within virtual communities, maintaining the same trust-based influence that underpinned the original model (Kosse and Carstens, 2023).

This shift in influence dynamics is supported by a growing body of empirical research. Social media influencers have been shown to play an active role in shaping political attitudes, influencing voting behaviours and triggering political mobilisation among their audiences (Bakshy et al., 2011; Betts and Bliuc, 2022; Borchers, 2025; Cheng et al., 2023; Flamino et al., 2023; Goodwin et al., 2023a, 2023b; Harff and Schmuck, 2023; Marriott, 2024; Prawira et al., 2024; Seeger and Muth, 2023; Soares et al., 2018).

Yet, despite their significance in empirical studies, influencers remain largely absent or simplistically represented in formal opinion dynamics models. This disconnect has created a blind spot in the field: models may describe consensus or polarisation within a network, but without accounting for the unique structural and psychological leverage of influencers, they risk missing out a key driver of how opinion actually forms and evolves in contemporary digital societies.

Within opinion dynamics, some work has been done on the effects of certain types of individuals including: charismatic leaders (Hegselmann and Krause, 2015), zealots (stubborn agents with staunch opinions) (Baumann, Sokolov, and Tyloo, 2020; Cao et al., 2024; Mobilia et al., 2007; Verma et al., 2014), mass media (Brooks and Porter, 2020; Hu et al., 2024) and highly influential agents (Chazelle, 2012; Coculescu et al., 2023; Helfmann et al., 2023). Most of these studies rely on socio-physical models rooted in numerical analysis, producing outcomes that, while inspired by socially relevant questions, are largely rooted in complex mathematical modelling. As a result, they remain inaccessible and of limited practical use to social scientists. Also, a number of these works utilise the voter model as a means to incorporate agent behaviour without either (1) incurring significant computational costs, or (2) oversimplifying inherently complex mathematical structures. As such, the results are not directly comparable to those derived from continuous models, like ours, despite addressing similarly relevant social dynamics.

Other studies have taken a mathematical approach to identifying optimal strategies for exerting influence over opinions. For instance, Hegselmann et al. (2015) presents a methodologically rigorous exploration of opinion manipulation through optimal control theory. However, their analysis is limited by the scale of the network, which consisted of only ten nodes. Additionally, the study's definition of an influencer lacks grounding in realistic social dynamics. Instead of reflecting observable behavioural traits or optimisation criteria, influencers were defined purely by their capacity to steer the opinion distribution toward a predefined target.

A closely related study within the field of innovation diffusion examines the role of special agents, referred to as influentials, in driving the adoption of ideas and innovations within agent-based networks (D. Watts and Dodds, 2007). This paper is notable for grounding its model firmly in empirically available evidence and for explicitly addressing a key gap in the literature regarding the impact of influentials. However, the diffusion model used relies on a binary adoption model. In contrast, our approach incorporates continuous opinion dynamics and iterative updating processes. As a result, the two models are not directly comparable in terms of outcomes, as they capture different facets of influence and decision-making.

Furthermore, the continuous model in Brooks and Porter (2020) has made significant strides in exploring the role of influential media accounts (influencers) in opinion dynamics. This study shares thematic similarities with ours, particularly in its focus on the impact of influencers on opinion formation and polarisation online. While the model offers valuable insights into echo chambers and incorporates important parameters such as information quality, it lacks dynamism. It thus fails to reflect the changing nature of relationships in real social networks. Additionally, its reliance on pairwise interactions may oversimplify the complex, large-scale connectivity typical of social media platforms. Despite these limitations, the study admirably attempts to bridge observed phenomena with its model. It also acts to guide data collection on online attitudes to validate its findings.

Also relevant, Helfmann et al. (2023) explored the influence of social media influencers and mass media on opinion dynamics, focusing on the macro-level effects of four influencers acting on a population of 250 nodes. Rather than employing an agent-based model, the study used a mean-field approach which, while computationally efficient, introduces two key limitations. First, the mathematical formalism of mean-field models sacrifices accessibility, making it difficult to establish strong empirical links or explain contextually clear mechanisms. Second, mean-field models typically obscure micro-level dynamics, limiting the ability to assess how specific influencer strategies affect individual users or their immediate network surroundings. This constrains the analysis of e.g. the types of agents that were being targeted.

Nonetheless, the study made valuable contributions by connecting its findings to relevant social science literature and offering actionable strategies to mitigate political polarisation online. The model's inclusion of several real-world digital components, such as recommendation algorithms and mass media influence, alongside social media influencers, was particularly commendable. However, combining all these factors into a single model diluted understanding their individual effects. A more modular approach, separating these agents, could have provided clearer insight into the distinct roles each plays in shaping opinion dynamics.

Despite their differences, we will reference all three studies later in our project. This is not only for comparative analysis, but also to draw on their frameworks to inform or defend our own modelling choices.

While our primary focus is on political online discourse, the insights we develop are applicable to any opinion-driven or polarised topic. We specifically choose online discourse to align our work with contemporary literature and increased empirical evidence in this domain. This enables us to justify our model design and insights by connecting to politically themed literature where applicable.

Our research is guided by a set of key questions that aim to deepen our understanding of the mechanisms and consequences of influencer-driven dynamics. While the existing literature has largely established that influencers can impact the system, less attention has been paid to how their influence unfolds within social systems, the boundaries of their impact, and the broader societal implications, particularly with regard to polarisation and extremism.

- What strategies do influencers employ to gain widespread influence, and how effective are these strategies? Are there limitations to their power?
- How does the population (or network) respond to influencers? Do we observe wide-spread consensus in opinion, or increased fragmentation and polarisation?
- Is the influencer's impact broad and diffuse, or more focused and targeted? What are the implications of this for the population?
- To what extent can our findings shed light on the drivers, mechanisms, and consequences of extremism and polarisation linked to online influencers?

Chapter 3

Base Model

In this chapter, we introduce the model of opinion dynamics, originally developed by Sayama (2020) and later extended by Bullock and Sayama (2023).

This model is chosen as a starting point for three main reasons. First, its core mechanisms, particularly the iterative opinion update rule, align closely with established continuous opinion models like Degroot (1974) and Hegselmann and Krause (2002). Second, its interaction dynamics are rooted in empirically validated concepts like homophily. Third, by adopting an agent-based modelling approach rather than a numerical method, we can conduct a micro-level analysis of interaction processes at the node level. This allows for a detailed understanding of how influencers can specifically impact the model's dynamics and the strategies they take to do so.

3.1 Structure

Let $G = (V, E)$ be a directed, weighted network, where V is the set of nodes, and E is the set of edges.

The network is fully connected, meaning for every pair of nodes (i, j), there is a directed edge from j to i and vice versa. The weight of the edge from node j to i is denoted by $w_{ij} \in \mathbb{R}$, and reflects the influence of j on i.

Each node $i \in V$ represents an individual, and each individual i holds an opinion or state $x_i \in \mathbb{R}$. This represents a stance or opinion on e.g. a political matter, such that an opinion of 1 may be less extreme than an opinion of 5. Similarly, an opinion of -1 may be more right wing and a +1 may be more left wing. Alternatively, the opinion may reflect a more specific viewpoint, such as support for open borders. In this case, an individual with an opinion of -1 would moderately oppose open borders, while a score of +1 would indicate moderate support for them.

The basic assumption of the model is that individuals shape their opinions towards their social norm, that is the weighted average opinion of an individual's local community. This principle, known as conformity, drives the evolution of individual opinions within the model. The edge weights and thus influence relationships evolve simultaneously, guided by two key principles: homophily and attention to novelty. Homophily refers to the tendency of similar individuals to strongly influence each-other. Attention to novelty reflects the human tendency to be influenced by those with extreme or unique opinions relative to your community. As a result, we have an adaptive network in which both individual opinion states and the edges between individuals co-evolve over time.

3.2 Adaptive and Evolutionary Processes

3.2.1 Opinion Update Mechanism

Opinion changes are governed by the principle of **conformity**. The change in the opinion x_i of node i over time is modelled by the following equation:

$$\frac{dx_i}{dt} = c\left(\langle x \rangle_i - x_i\right) + \epsilon \tag{3.1}$$

Here, c represents the parameter that defines the strength of **social conformity**.

The term $\langle x \rangle_i$ represents **the local social norm (or just social norm)** as perceived by node i and is computed as:

$$\langle x \rangle_i = \frac{\sum_{j \in N_i} w_{ij} x_j}{\sum_{j \in N_i} w_{ij}} \tag{3.2}$$

where N_i denotes the set of neighbours of node i, and w_{ij} represents the influence of node j on node i. Hence, the social norm is the weighted average of the opinions of node i's neighbours, where more influential neighbours have a greater impact on the value of $\langle x \rangle_i$. This means that node i's opinion is primarily shaped by the neighbours with the greatest influence over it. Finally, the parameter ϵ accounts for random fluctuations, representing various factors such as noise, external influences like media, inherent biases and more generally, the unpredictability of human behaviour.

The parameter c dictates how conformist each individual is, that is, the rate at which an individual is inclined to align with their community's opinion. The term $(\langle x \rangle_i - x_i)$ represents the difference between the local social norm and the individual's opinion, causing x_i to shift towards $\langle x \rangle_i$ at a rate controlled by the conformity parameter c. Intuitively, every individual moves towards their local social norm.

3.2.2 Relationships and Influence

The influence, or edge weights, evolve over time according to:

$$\frac{dw_{ij}}{dt} = hF_h(x_i, x_j) + aF_a(\langle x \rangle_i, x_j) \tag{3.3}$$

where h and a are parameters that control the strength of **homophily** and **attention to novelty**, respectively. The function $F_h(x_i, x_j)$ quantifies the similarity between opinions, reinforcing connections between like-minded individuals. In contrast, $F_a(\langle x \rangle_i, x_j)$ quantifies how different an opinion is from the social norm, capturing the influence of unique perspectives. Together, these factors determine how the influence between nodes evolves over time.

3.2.3 Homophily

The principle of homophily refers to the tendency of individuals to form stronger connections with those who share similar opinions (McPherson et al., 2001; Sayama, 2020). As a result, the more similar two individuals are, the greater their influence on each other. In the model, this is captured by the function $F_h(x_i, x_j)$:

$$F_h(x_i, x_j) = \theta_h - |x_i - x_j| \tag{3.4}$$

where θ_h is a threshold parameter that controls how much similarity in opinion is needed for individuals to positively or negatively influence each other. Looking at Eq. 3.7, if we keep F_a constant then we can consider how w_{ij} changes wrt. homophily/F_h. When the difference $|x_i - x_j|$ is smaller than θ_h, homophily is amplified, strengthening the influence of j on i. . Conversely, when $|x_i - x_j|$ exceeds θ_h, the individuals' opinions are too dissimilar, thereby weakening the influence of j on i.

For larger values of h, the increase in influence will be greater, while for smaller values of h, the influence will change more gradually. This parameter, therefore, controls the rate at which a the influence between two nodes increases or decreases.

3.2.4 Attention to Novelty

Attention to novelty refers to the tendency of humans to be influenced by extreme or unique opinions (Goldenberg et al., 2023; Ramachandran, 2010). This is modelled by the function $F_a(\langle x \rangle_i, x_j)$:

$$F_a(\langle x \rangle_i, x_j) = |\langle x \rangle_i - x_j| - \theta_a \tag{3.5}$$

Here, θ_a is a threshold parameter that controls how extreme and unique an opinion has to be in comparison to the local social norm of node i. When the difference $|\langle x \rangle_i - x_j|$ is greater than θ_a, the influence of j on i increases, reflecting an attraction to unique and extreme opinions. By using local social norm, we reiterate the notion that an individual's opinion is heavily swayed by their local community, particularly those closely connected to each-other.

If we look at Eq. 3.7 again, we can keep F_h constant then consider how w_{ij} changes wrt. attention to novelty/F_a. For a node i, if another node j has a very different opinion from i's social norm, it is likely that $|\langle x\rangle_i - x_j| \geq \theta_a$. This increases w_{ij}, strengthening the influence of j on i. If instead $|\langle x\rangle_i - x_j| \leq \theta_a$, the influence of j on i decreases.

Like homophily, the parameter a controls the rate at which w_{ij} is affected by attention to novelty / F_a.

3.3 Heterogeneous Parameters

We introduce heterogeneity to our model adapted from Bullock and Sayama (2023). This allows us to assign each node its own parameter values for c, h, and a. The variation in parameters simulates individual personality traits, introducing an additional layer of complexity to the model. The heterogeneous parameters were explored using both uniform and non-uniform (normal) distributions. However, the non-uniform distribution was implemented in such a way that they aimed to get a fragmented network of opinions. Since our goal is not to target a specific outcome, we opted for uniform heterogeneity, which better suited the studys objectives and realism.

We achieve this by drawing values for c_i, h_i, and a_i from a uniform distribution in the range $[0.01, 0.3]$, while keeping the values of θ_h and θ_a fixed:

$$c_i, h_i, a_i \sim \mathcal{U}(0.01, 0.3). \tag{3.6}$$

The equations governing edge weights and opinion updates become:

$$\frac{dw_{ij}}{dt} = h_i F_h(x_i, x_j) + a_i F_a(\langle x\rangle_i, x_j), \tag{3.7}$$

$$\frac{dx_i}{dt} = c_i\left(\langle x\rangle_i - x_i\right) + \epsilon. \tag{3.8}$$

Therefore, h_i affects an individual's tolerance of opinion differences whereas a_i controls an individual's appetite for bonding with more extreme individuals.

A detailed explanation of parametric effects on opinion dynamics is laid out below.

3.3.1 Discussion on Individual Parameters

Parameter c_i (Conformity) From Eq. 3.8, we see that c_i controls how strongly an individual's opinion x_i moves towards the social norm $\langle x\rangle_i$:

$$\frac{dx_i}{dt} = c_i\left(\langle x\rangle_i - x_i\right) + \epsilon.$$

A higher value of c_i means that node i will adjust their opinion more rapidly towards the collective opinion of their neighbours, making them more conformist. Conversely, a lower c_i implies that the node is more resistant to changing, and their opinion evolves more slowly towards the social norm, reflecting a more independent or non-conformist behaviour.

Parameter h_i (Homophily) The parameter h_i influences how strongly node i is affected by individuals with similar opinions.

Using Eq. 3.7, if we keep $a_i F_a(\langle x\rangle_i, x_j)$ constant:

- **If $F_h(x_i, x_j)$ is negative (i.e., $|x_i - x_j| > \theta_h$): The nodes have different opinions.** When h_i is large, the negative value of $F_h(x_i, x_j)$ is amplified, meaning the influence of j on i becomes weaker. The larger h_i is, the more node i will avoid being influenced by others with different opinions, reinforcing the homophily effect.

  On the other hand, when h_i is small, the influence of negative $F_h(x_i, x_j)$ is lessened, and node i is thus more tolerant of differences in opinions, resulting in a weaker negative effect on the relationship.

- **If $F_h(x_i, x_j)$ is positive (i.e., $|x_i - x_j| < \theta_h$): The nodes are more like-minded.** When h_i is large, the positive value of $F_h(x_i, x_j)$ is amplified, leading to a stronger influence from node j to i as their opinions align more closely. Larger h_i thus causes node i to be more influenced by individuals of similar opinion. Conversely, when h_i is small, the positive effect of $F_h(x_i, x_j)$ is reduced and the relationship is only marginally increased. Thus, node i is less sensitive to the influence of those with similar opinions.

Parameter a_i (Attention to novelty) The parameter a_i determines how sensitive node i is to extreme opinions or outliers in relation to a perceived social norm.

Using Eq. 3.7, if we keep $h_i F_h(x_i, x_j)$ constant:

- **$F_a(\langle x \rangle_i, x_j)$ is negative (i.e., $|\langle x \rangle_i - x_j| < \theta_a$): The node j has a similar opinion to node i's social norm.** When a_i is large, the negative value of $F_a(\langle x \rangle_i, x_j)$ is amplified, leading to a weaker connection between nodes i and j. In this case, node i is less influenced by nodes similar to the perceived social norm.

  Otherwise, when a_i is small, the influence of $F_a(\langle x \rangle_i, x_j)$ is reduced, meaning there is less of an effect on the influence relationship.

- **If $F_a(\langle x \rangle_i, x_j)$ is positive (i.e., $|\langle x \rangle_i - x_j| > \theta_a$): The node, j, has an unique opinion compared to node i's social norm.** When a_i is large, the positive value of $F_a(\langle x \rangle_i, x_j)$ is amplified, meaning the influence from node j increases as its opinion deviates more from the social norm. The larger a_i, the more node i will be influenced by these extreme individuals. When a_i is small, the positive effect of $F_a(\langle x \rangle_i, x_j)$ is reduced, and node i is a little more resistant to extreme opinions. This lessens the increase of j's influence when its opinion is significantly different from i's local average. Even though j is unique, i isn't too bothered.

3.4 Technical Implementation

This section outlines the implementation and configuration of each network and simulation. The Data Collation and Analysis section specifically describes how we structured our experiments, including the number of simulations and network runs.

Following the approach in Bullock and Sayama (2023), we implement an adaptive network model in Python 3.7 using the NetworkX library. Python's ease-of-use and quick development cycle, supported by a wide range of libraries make it an ideal choice for simulating complex, evolving systems. NetworkX, in particular, is an efficient and widely used toolkit for graph-based modelling, offering several inbuilt tools to handle complex graph structures central to our project.

The model consists of an initial set-up and a simulation process. We detail these below.

Network Set-Up

1. **Initialise Network Object:** We begin by initialising a fully-connected graph object with 1000 nodes using NetworkX.

2. **Set Parameters:** : The values of c_i, h_i, and a_i are drawn from a uniform distribution in the range $[0.01, 0.3]$, while the threshold values θ_h and θ_a are fixed at $\theta_h = 0.3$ and $\theta_a = 0.3$, respectively.

$$c_i, h_i, a_i \sim \mathcal{U}(0.01, 0.3). \tag{3.9}$$

3. **Set Starting Opinions:** For each node, a real value is assigned from a normal distribution $\mathcal{N}(0, 1)$ to represent its opinion.

4. **Set Starting Edge Weights:** Each edge is assigned a random real value from the interval $[0, 1]$, ensuring that all influence relationships are positive and less than 1 at the start.

Simulation Process

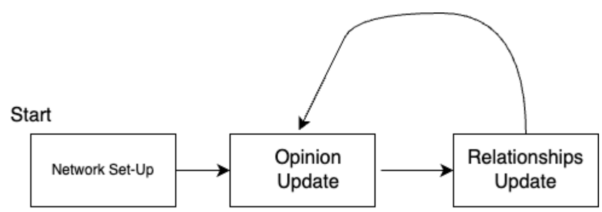


Figure 3.1: Illustration of the set-up and simulation process.

A single simulation is conducted over 1000 time steps, where at each time step, the following updates are made. We note that the order of the nodes is shuffled randomly at each time step to prevent synchronicity effects.

1. **Opinion Update**: The state of the nodes is updated according to Eq. 3.8. The small fluctuation, ϵ, is drawn from a normal distribution with mean 0 and standard deviation 0.01:

$$\epsilon \sim \mathcal{N}(0, 0.01)$$

2. **Relationships Update**: The edge weights are updated using Eq. 3.7. Any negative weights are set to zero to maintain positivity in the system.

The equations for the updates in Step (1) and (2) are multiplied by the time step, dt, corresponding to 0.1. Figure 3.1 illustrates this process.

3.5 Outcome Measures

We use the same outcome measures as in Bullock and Sayama (2023). We will refer to these later as the Base Model Outcome Measures in Chapter 4.

1. Mean Edge Weight
2. Assortativity
3. Community Size Standard Deviation
4. Network Modularity
5. Community Mean Opinion Standard Deviation
6. Mean Node Deviance

While several attempts have been made to ensure consistency in aggregating the original measures during calculations, discrepancies may arise due to the lack of clear definitions in Bullock and Sayama (2023). Similarly, early in this project, we observed discrepancies in the calculation of communities. This likely stems from our use of NetworkX's built-in Louvain algorithm, whereas the original paper used its own. As a result, minor numerical differences in the values for Outcome Measures 3 and 5 should be expected but the overall patterns should be the same.

To best understand these measures, it may be useful to review the concept of a **community** as clarified in Chapter 2. We now provide some specific definitions of our own measures:

Outcome Measure 1 - Mean Edge Weight

Mean edge weight measures the average value of w_{ij} between nodes in a network, and is mathematically defined as:

$$\text{Mean Edge Weight} = \frac{1}{|E|} \sum_{(i,j)\in E} w_{ij}$$

Intuitively, this is the average influence between two individuals. It provides an idea on how what the average influence relationship is in the population and is a baseline for comparing such relationships.

Outcome Measure 2 - Assortativity

Assortativity measures the tendency of nodes to connect with those that share similar attributes or characteristics. Since this measure was originally defined with respect to unweighted, undirected graphs, we have adapted the method developed by Yuan et al. (2021) for our graph.

Intuitively, assortativity quantifies the significance of homophily within the network. Since it reveals how likely nodes are to form connections with similarly-minded others, it provides a measure of how strongly homophily influences the structure of the network. This helps to understand how tightly-knit nodes are in relation to opinions.

Unless specified otherwise, we measure assortativity in relation to opinions. This means we assess the nodes on their tendency to connect with others who share a similar opinion value.

Outcome Measure 3 and 5 - Community Size Standard Deviation and Mean Opinion Standard Deviation

Community size is the number of nodes or individuals within a community. For each network, the standard deviation of community sizes is defined as:

$$\sigma_{\text{size}} = \sqrt{\frac{1}{K}\sum_{k=1}^{K}(C_k - \mu_C)^2}$$

where C_k is the size of community k, μ_C is the mean community size, and K is the total number of communities.

Community opinion standard deviation measures the standard deviation of opinions within a community. For each community k, the opinion standard deviation is defined as:

$$\sigma_{\text{opinion},k} = \sqrt{\frac{1}{|C_k|}\sum_{i \in C_k}(x_i - \mu_x)^2}$$

where x_i is the opinion of node i in community C_k, μ_x is the mean opinion within community C_k, and $|C_k|$ is the number of nodes in community C_k. This gives us an idea of how spread out opinions are within each community. A low value would suggest most communities are formed of very like-minded individuals, and vice-versa for higher values.

Outcome Measure 4 - Network Modularity

Modularity allows us to assess whether a network is effectively partitioned into communities. We gave a specific definition of this in Chapter 2. Low modularity indicates that the nodes within a community are not significantly more connected to each other than to those outside the community. Likewise, high modularity signals that the nodes in a community are strongly connected within the group and weakly related to nodes outside the group.

Contextually, modularity can be considered as a measure of how well a social network is divided into closely bonded groups, reflecting stronger interpersonal connections within each community than to non-community members in the network.

Outcome Measure 6 - Node Deviance

Node deviance is the absolute difference between a node's opinion and its network's average. Mathematically, it is defined as:

$$\text{Node Deviance}_i = |x_i - \mu_o|$$

where x_i is the opinion of node i and μ_o is the average opinion of the entire network.

Intuitively, the mean node deviance of a network tells us how spread out the opinions in the population are. Low node deviance implies a more similar population, whereas high node deviance implies a higher proportion of individuals with unique opinions.

3.6 Data Collation and Analysis

As detailed in the Technical Implementation section, the specific technical configuration for each network is provided. In this section, we focus on the experimental setup, outlining the configuration of the simulations and the process for data collection and analysis. The results of which are in Chapter 5.

Recall that heterogeneous networks are those with heterogeneous c, h and a whilst homogeneous networks use fixed values.

Experimental Set-Up

1. We simulate 5 **heterogeneous** networks of n = 1000 for 1000 time-steps.
2. We then take the average c_i, h_i and a_i values for each **heterogeneous** network and use these as our fixed values for c, h and a in the set up of our **homogeneous** networks.

3. We then simulate these 5 **homogeneous** networks for 1000 time steps.

At the end of all simulations, we have a set of ten networks: five homogeneous and five heterogeneous. This ensures a fair comparison between the networks.

Data Collation

Simulations are executed on the University of Bristol's BluePebble supercomputer which is accessed remotely via (SSH) from a local machine. By using the sbatch command, multiple simulations can be parallelised, improving computational efficiency.

For each network, a corresponding pickle file containing its NetworkX graph object is generated. Consequently, for both homogeneous and heterogeneous networks, a total of ten pickle files are produced.

Pickle is employed due to its high efficiency in serialising and deserialising complex data structures, such as NetworkX graph objects, while maintaining their structure. Given the large-scale nature of the simulations run on the supercomputer, using Pickle allows for the effective storage and retrieval of extensive network data between machines. This approach reduces computational overhead because serialised files can be easily transferred from the supercomputer to local storage.

Statistical Measures

Where statistical tests are mentioned:

- For each measure, Brown-Forsythe tests were conducted to check for equality of variance.
- Depending on this output, a two-sample t-test is performed to identify statistically significant differences between our observations. This helps ascertain if the two observations e.g. edge weights, share similar statistical properties.
- Where effect size is mentioned, we use Cohen's D-value to ascertain the scale of difference between two datasets (relative to their means). In this project, our focus is on the comparative scale of these differences rather than the absolute D-values themselves.
- Where correlation analysis is used, we use Pearson's correlation coefficient to measure the linear relationship between two variables. This determines the strength of the association between the two.

These tests were performed primarily using the SciPy stats library, as it provides an efficient set of statistical functions that are widely used and well-documented.

Chapter 4

Extended Model

We extend this base model by incorporating two new features: an attention limit and influencer agents.

To explore the role of influence in shaping network opinions, we define two distinct scenarios: single influencer and multiple influencers. This provides a broader scope, allowing us to compare local versus global influence mechanisms and outcomes.

4.1 Attention Limit

An **attention limit** is used to model individuals' limited tolerance for highly divergent opinions. This reflects the well-documented psychological phenomenon of confirmation bias, where people tend to seek out, interpret, and remember information that aligns with their pre-existing beliefs (Nickerson, 1998). Individuals actively shape their informational environments by selectively engaging with sources and individuals who reinforce their views (Festinger, 1957; Frey, 1986; Klapper, 1960). This tendency has been extensively studied in psychology and cognitive science (Bacon, 1620; Doherty and Mynatt, 1987; Festinger, 1957; Frey, 1986; Klapper, 1960).

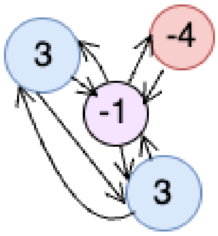


Figure 4.1: Unweighted network of 4 nodes with varying continuous opinions. Bonds are formed between individuals whose opinions are within 5-points, such as between nodes with opinions of 3 and -1, while no bonds are formed between nodes with opinions of -4 and 3 due to their ideological distance.

Formal Definition

We implement this limit using the following:

If the opinion difference between x_i and x_j is greater than or equal to 5, we set the weight $w_{ij} = 0$, such that their influence and thus relationship is cut off. This condition is checked when the edge between two individuals is evaluated.

Formally:

$$w_{ij} = \begin{cases} 0, & \text{if } |x_i - x_j| \geq 5 \\ w_{ij}, & \text{otherwise} \end{cases} \tag{4.1}$$

Similarly, the condition applies symmetrically to w_{ji}:

$$w_{ji} = \begin{cases} 0, & \text{if } |x_j - x_i| \geq 5 \\ w_{ji}, & \text{otherwise} \end{cases} \tag{4.2}$$

Figure 4.1 illustrates how ideologically far apart nodes will not influence each other.

The concept of an attention limit is analogous to the thresholds in bounded-confidence models, such as the original H-K model (Hegselmann and Krause, 2002). They are also both driven by the same sociological roots. However, we avoid using this terminology, as our project diverges from the methodologies typically associated under bounded-confidence models.

4.2 Influencers

Crucial to this research is the incorporation of a special type of agent that we call **influencers**. These agents are randomly selected individuals who, at every time step, skip the typical opinion dynamic rules of their neighbours and evolve according to their own strategy. This strategy chooses opinions that maximise the influencer's total influence on all its neighbours. This is synonymous with an individual meticulously designing their social media content when trying to gain followers online. Unlike the average person, these influencers are tactical about what they post in a bid to gain a large online presence. The strategy of using targeted content to gain followers and build an online presence is well-documented in literature (Borchers, 2025).

4.2.1 Multiple Influencers

Finally, we investigate the impact of having 10% of the population act as influencers within the network, as opposed to a single influencer. This extends previous work on influencers, such as Helfmann et al. (2023), which focused on the effect of a lone influencer, and builds upon the findings in Brooks and Porter (2020), which explored the dynamics of multiple influencers within a network. While our study only considers a fixed percentage of 10%, other research has examined the effects of varying this number as in Brooks and Porter (2020).

These agents operate the same as our influencers but we simulate the two scenarios independently for a separate analysis.

4.3 Influencer Design

4.3.1 Incorporating Influencers into the Network

We start by randomly choosing one node from the set of all nodes V, to be our influencer.

We allow the rest of the network to be set-up exactly the same as before, but on every time step, the influencer activates a special function. Importantly, this occurs before the rest of the network updates within the same time step.

The basic idea behind the function is that an influencer will assess the possible opinions it could take, and chooses the opinion which will maximise its total influence on others. Formally, it maximises the value of its out-strength, as in Eq. 2.2.6.

4.3.2 The Behaviour of Influencers

Our initial network set-up is identical to our base model. We now detail the strategy used by our influencers in detail.

Let the set of possible opinion shifts be defined as:

$$\Delta = \{-1, -0.9, \ldots, 0.9, 1\}$$

On the **first** time step, the influencer adopts the opinion of its local social norm, as described in Eq. 3.2. It essentially disregards its own true opinion and aligns with the average opinion within its neighbourhood.

On subsequent time steps, the influencer tests each candidate opinion, defined as its current opinion plus each value in Δ, or $x_{\text{inf}} + \Delta$. For each candidate opinion, it calculates the resulting change in influence relationships with all its neighbours. Cumulatively, this forms the out-strength, as defined in Eq. 2.2.6. The best opinion is the one that results in the greatest increase in total out-strength for the influencer.

This opinion then overrides the influencer's 'true' opinion and the simulation continues as normal for the other agents. This influencer function effectively happens before Step (1), the network-wide opinion update, in the simulation process detailed in Chapter 3. Therefore, on every time-step, the influencer gets to strategically adopt a new opinion before the rest of the network updates as in Figure 4.2.

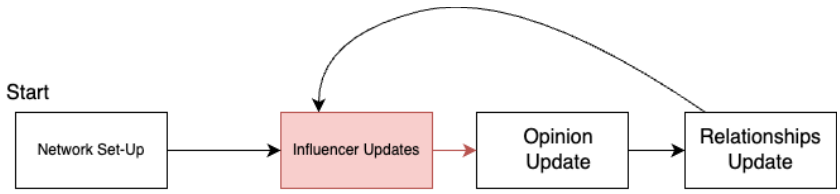


Figure 4.2: Simulation process after the influencers are integrated into the model. The influencers update comes before the network-wide update for both opinions and relationships.

Formal Definition

Let the influencer be node inf and the test opinion be $\hat{x}_{\text{inf}}$, where $\hat{x}_{\text{inf}} = x_{\text{inf}} + \Delta$.

For each neighbour j of inf, the change in the total out-strength is calculated by:

$$\Delta w_{j,\text{inf}} = \Delta w^{\text{a}}_{j,\text{inf}} + \Delta w^{\text{h}}_{j,\text{inf}}$$

where $\Delta w^{\text{a}}_{j,\text{inf}}$ is the change between two individuals due to attention to novelty and $\Delta w^{\text{h}}_{j,\text{inf}}$ is the change between two individuals due to homophily.

The influencer selects the opinion $\hat{x}_{\text{inf}}$ that maximises the total new out-strength, denoted by $\text{OS}^{\text{new}}_{\text{inf}}$, where:

$$\text{OS}^{\text{new}}_{\text{inf}} = \sum_{j \in \text{Neighbours(inf)}} \left(w_{j,\text{inf}} + \Delta w^{\text{a}}_{j,\text{inf}} + \Delta w^{\text{h}}_{j,\text{inf}} \right)$$

where $w_{j,\text{inf}}$ is the current weight of the edge from inf to j, and $\Delta w^{\text{a}}_{j,\text{inf}}$ and $\Delta w^{\text{h}}_{j,\text{inf}}$ capture changes due to attention to novelty and homophily, respectively. The optimal opinion choice is then:

$$\hat{x}^{*}_{\text{inf}} = \arg\max_{\hat{x}_{\text{inf}}} OS^{\text{new}}_{\text{inf}}$$

ensuring that the influencer adopts the opinion that results in the highest possible out-strength.

4.3.3 Summary of Influencer Behaviour

In summary:

1. At the first time-step, the influencer adopts the local social norm.
2. The influencer then tests a range of candidate opinions, evaluating the resulting change in influence for each neighbour.
3. The influencer then selects the opinion that maximises the changes in influence across all its neighbours, overriding its true opinion.

This process is repeated on each time-step throughout the simulation process. This means the influencer is never subject to the same update mechanism, Eq. 3.8, as the other agents in Step (1) of the simulation process.

The strategy adopted reflects the defining traits of influencers, namely, their goal to achieve greater following and thus influence over their peers. While we initially considered a continuous optimisation approach, it proved computationally expensive at scale. Instead, we use discrete opinion shifts, which offer key benefits: they provide strategic flexibility for influencers to adopt more extreme or moderate positions as needed; they allow for gradual, realistic changes that align with how influencers maintain brand consistency; and they significantly reduce computational overhead, making the model more scalable. This scalability is essential in agent-based modelling, especially when incorporating exogenous agents such as influencers, whose inclusion could otherwise introduce significant time complexity in the simulation process.

4.4 Multiple Influencer Design

For multiple influencers, we randomly select 10% of the nodes in our network to act as influencers. These agents evolve according to the same dynamics as described in the single influencer model. The choice of this percentage was driven by computational constraints; balancing simulation time without compromising network size, but other works have varied this number (D. Watts and Dodds, 2007).

While single influencer simulations are primarily used to understand influencer-specific dynamics, the multiple influencer framework shifts focus to the broader impact on network structure. This design allows us to assess how the presence of multiple, distributed sources of influence shapes global network outcomes, such as consensus and polarisation.

4.5 Outcome Measures

4.5.1 Influencer Behaviour

Opinion and Out-Strength Evolutions

We record the opinions and total out-strengths of the influencer nodes at regular intervals (every ten time steps) to provide a comprehensive understanding of how our influencers evolve over time. Details of this are provided in the Data Collation and Analysis section below.

4.5.2 Top 50 Neighbours

An in-depth analysis is conducted on the top 50 neighbour nodes who experience the greatest influence from the influencer. Specifically, the top 50 neighbours are determined by selecting the nodes with the highest edge weights from the influencer, as given by the following procedure:

$$\text{Top 50 Neighbours} = \text{sorted}\left(\{(nbr, w_{ij}) \mid nbr \in \mathcal{N}(\text{influencer})\},\ \text{key} = \lambda x : x[1],\ \text{reverse} = \text{True}\right)[: 50]$$

where $\mathcal{N}(\text{influencer})$ denotes the set of all direct neighbours of the influencer, and w_{ij} is the weight of the directed edge from the influencer i to the neighbour j. Sorting is done in descending order of w_{ij} to prioritise stronger connections. Only the top 50 nodes are chosen to focus on the upper tail of the influence distribution (the neighbours receiving the greatest influence). This threshold addresses two needs: (1) isolating the strongest influence relationships, and (2) maintaining a large enough sample for an in-depth statistical analysis without diluting the results with more weakly influenced nodes.

By choosing this subset of neighbours, we can better evaluate the micro-level details of the influencers strategies. Whilst various visuals are illustrated in Chapter 6, we clearly define the following measures:

Opinion Distance from Network Average/to Influencer

The opinion distance, D, from the network average is defined as the average distance between a neighbour's opinion x_i and the network's average opinion $\langle x \rangle$:

$$D_{\text{avg}} = \frac{1}{N} \sum_{i=1}^{N} |x_i - \langle x \rangle|$$

Similarly, the opinion distance, D, to the influencer is the average distance between a neighbour's opinion x_i and the influencer's extreme opinion x_{inf}:

$$D_{\text{inf}} = \frac{1}{N} \sum_{i=1}^{N} |x_i - x_{\text{inf}}|$$

In both cases, D represents the distance metric used to assess: the extremity of a nodes opinion and how close their opinion is to the influencer. For our measures, we calculate the average D value per network, then across all networks.

4.6 Data Collation and Analysis

As outlined in the Technical Implementation section of Chapter 3, the network setup remains the same. The sole modification lies in the simulation process, where influencer agents strategically adopt an opinion prior to the opinion and relationship updates (Steps 1 and 2).

In this section, we focus on the experimental setup, outlining the configuration of the simulations and the process for data collection and analysis.

Experimental Set-Up

In each run, we generate two graphs, G and H, which have the exact same initial conditions. Formally, we have:

$$G = (V, E) \quad \text{and} \quad H = (V, E).$$

- For all nodes $i \in V$, their initial opinions are identical:

$$x_i^G(0) = x_i^H(0).$$

- For all edges $(i, j) \in E$, the edge weights are the same in both graphs:

$$w_{ij}^G = w_{ij}^H.$$

We randomly select a node to act as the influencer. In graph G, this agent engages in influencer interactions. Graph H will act as a null model in which we can compare the impact of the influencer node.

We perform 30 simulations, resulting in 60 networks—30 instances of G graphs and 30 instances of H graphs. Each network is simulated for 1000 time-steps with n = 1000 nodes in each simulation.

Data Collation

All simulations, along with their corresponding pickle files, are collated as described in Chapter 3. However, to track key metrics associated with the influencer node's opinion and out-strength, as well as temporal measures such as the network's opinion dynamics, mean out-weight, and others, the results are recorded in a CSV file.

The following data is saved for both graphs G and H, with results separated by network type:

- **Time**
- **Mean Opinion** (average network opinion)
- **Influencer Opinion**
- **Local Social Norm of Influencer**
- **Influencer In/Out Weights**
- **Average Network In/Out Weights**
- **Network Type** (influencer or non-influencer/G or H)

To facilitate this process, a CSV file is created and updated with each simulation run. The relevant data is appended to the file, with headers written if the file is newly created. Upon completion of the simulations, these CSV files can be transferred to local storage for further analysis, ensuring that all simulation results are consolidated and accessible for in-depth examination.

Due to the exogenous nature of the influencers, simulating multiple influencers results in significantly higher time expenses compared to simulating a single influencer. As a result, we perform only a single simulation, generating both the G and H graphs for a scenario with and without influencers.

Statistical Measures

Similar to data collation, all statistical measures are those described in Chapter 3.

Chapter 5

Base Model - Results

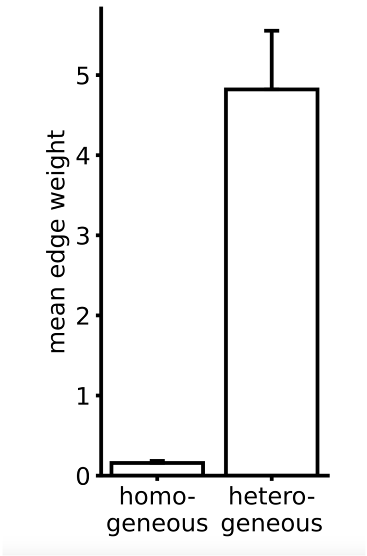
In this chapter, we examine the results of our base model using outcome measures derived from the original paper (Bullock and Sayama, 2023). The aim is to replicate these results to ensure our model has a solid foundation from which to extend from.

5.1 Outcome Measures

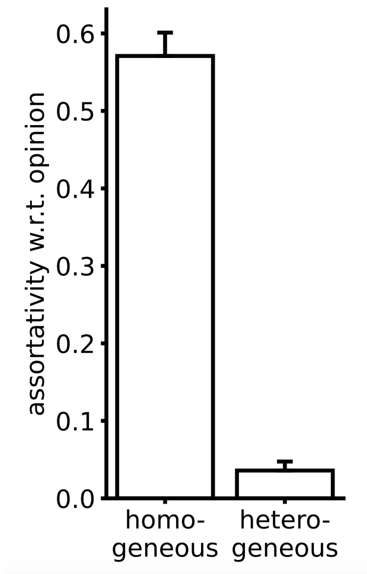
Figures 5.1 and 5.2 compare the results from the original study (Bullock and Sayama, 2023) with those obtained from our own simulations. The result is a comparison of bar charts for each different outcome measure detailed in Chapter 3 with similarities across all metrics.

As we mentioned in Chapter 3, we anticipated differences in some of our community measures, see Figures 5.2b, 5.2a, 5.2d, 5.2e. Whilst the underlying pattern is consistent between our studies, numerical differences can be attributed to (1) poor detailing of the measures in the original paper and (2) a different Louvain-method algorithm.

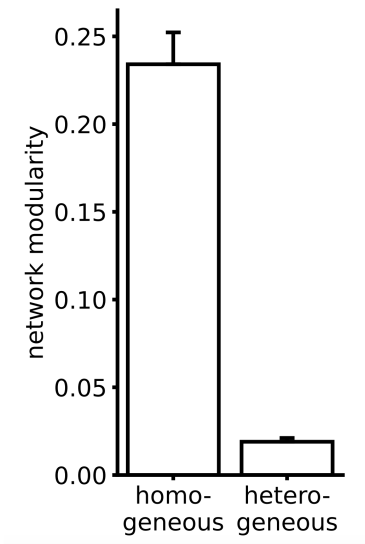
All comparisons within our own simulations yielded p-values below 1×10^{-11}, confirming statistically significant differences consistent with the original paper's findings. Table A.4 in Appendix A shows the full set of statistical results for these metrics.



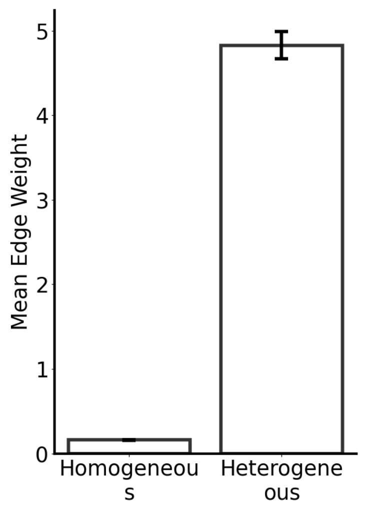
(a) Original Study: Mean Edge Weight



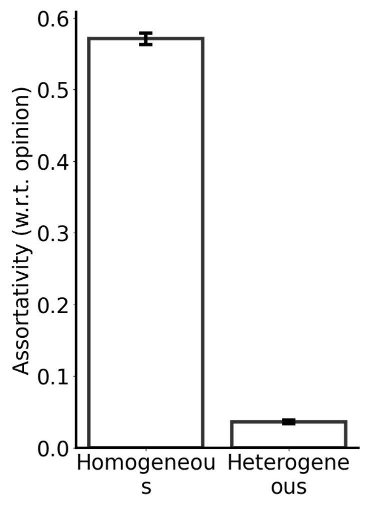
(b) Original Study: Assortativity



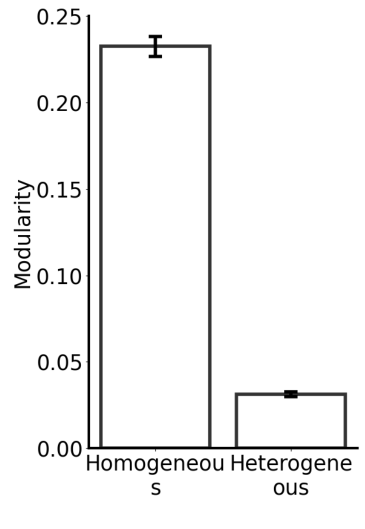
(c) Original Study: Modularity



(d) Our Simulations: Mean Edge Weight

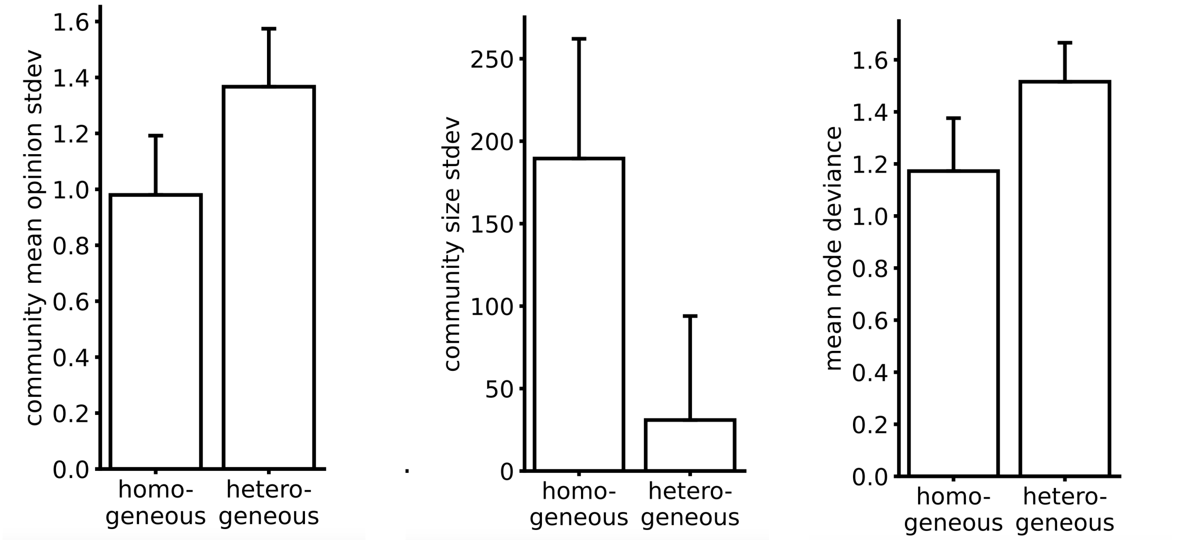


(e) Our Simulations: Assortativity



(f) Our Simulations: Modularity

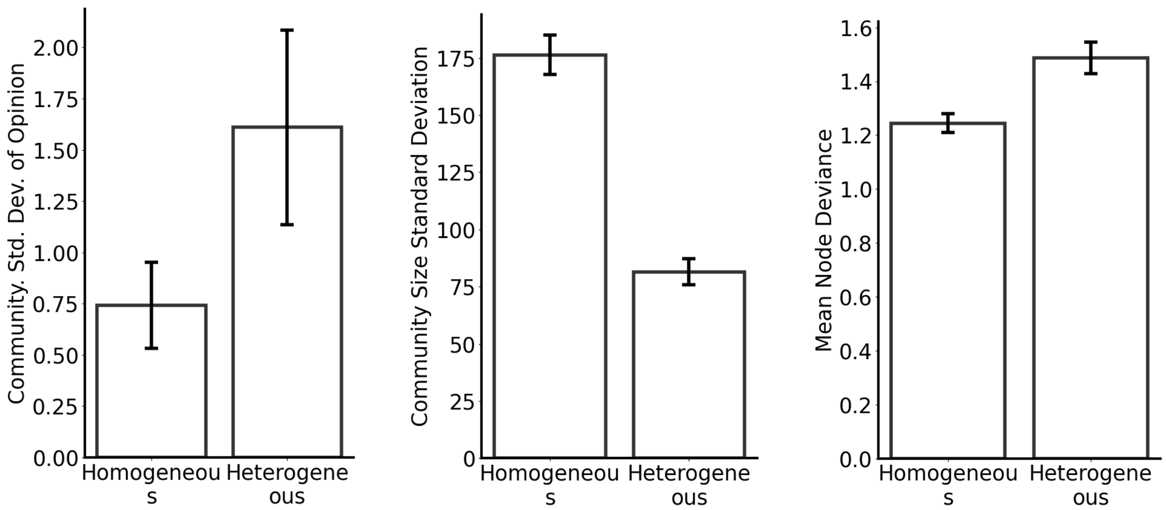
Figure 5.1: Side-by-side comparison of individual outcome measures between original study and our own simulations of heterogeneous and homogeneous networks ($n = 1000$).



(a) Original Study: Community Opinion Std. Dev.

(b) Original Study: Community Size Std. Dev.

(c) Original Study: Mean Node Deviance



(d) Our Simulations: Community Opinion Std. Dev.

(e) Our Simulations: Community Size Std. Dev.

(f) Our Simulations: Mean Node Deviance

Figure 5.2: Side-by-side comparison of individual outcome measures between original (top) and new (bottom) simulations of heterogeneous and homogeneous networks ($n = 1000$).

Chapter 6

Extended Model - Results

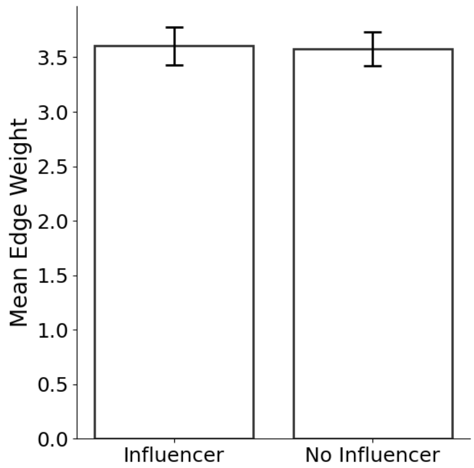
This section illustrates visuals to address some of our key questions from Chapter 2.

- What strategies do influencers employ to gain widespread influence, and how effective are these strategies? Are there limitations to their power?
- How does the population (or network) respond to influencers? Do we observe wide-spread consensus in opinion, or increased fragmentation and polarisation?
- Is the influencers impact broad and diffuse, or more focused and targeted? What are the implications of this for the population?

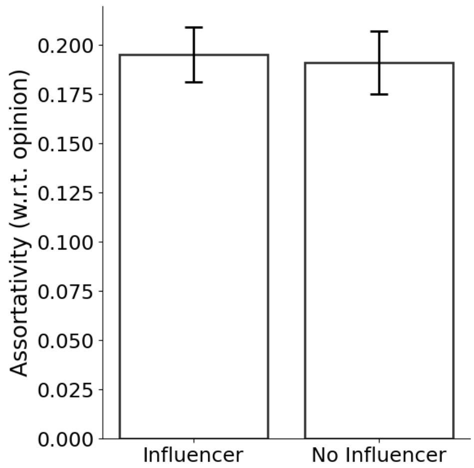
6.1 Base Model Outcome Measures

In Figure 6.1, we present the outcome measures derived from Bullock and Sayama (2023) to compare influencer and non-influencer networks. These measures serve as critical points of reference for assessing changes in the overall network structure. For all metrics, with the exception of community size standard deviation, two-sample t-tests revealed no statistically significant differences between influencer and non-influencer networks ($p > 0.05$). While community size standard deviation did exhibit a statistically significant difference, the effect size (Cohen's d = -0.245) suggests that the observed difference is of minimal practical significance within the context of this project. This likely stems from ambiguities in the original study and differences in methodology, specifically, their manual implementation of the Louvain method versus our use of the built-in NetworkX function.

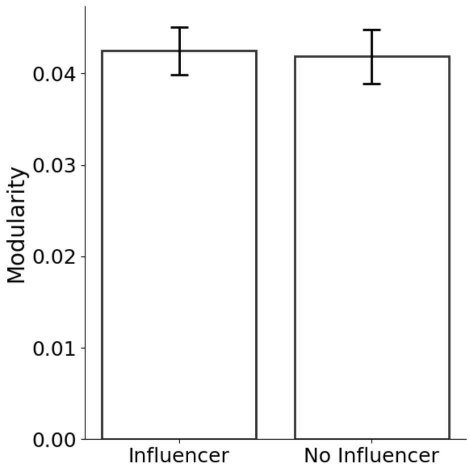
Refer to Table A.1 in Appendix A for a comprehensive summary of the statistical analyses conducted on measures against influencer and non-influencer networks.



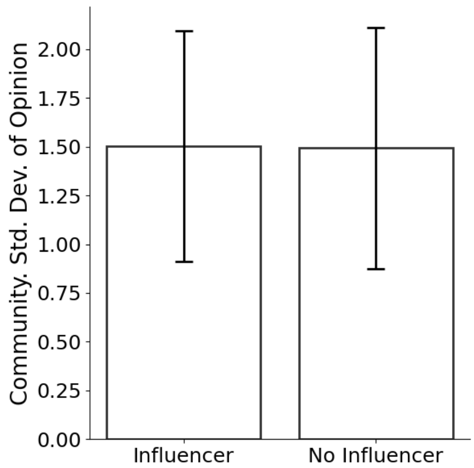
(a) Mean Edge Weight



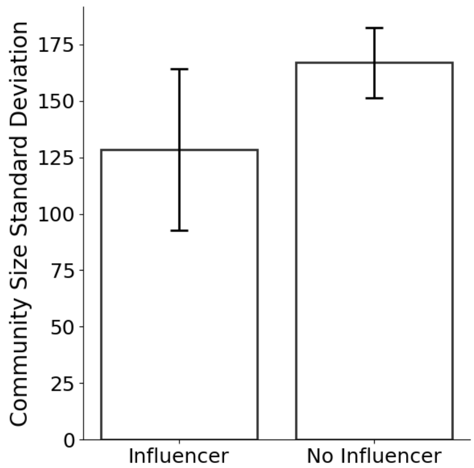
(b) Assortativity



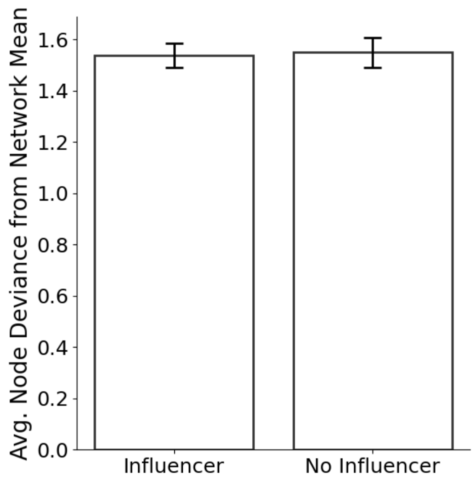
(c) Modularity



(d) Standard Deviation of Community Opinion



(e) Community Size



(f) Mean Node Deviance

Figure 6.1: Comparison of outcome measures in non-influencer and influencer networks. The results show similar patterns for both network types, indicating that the presence of influencers does not significantly alter the overall network outcome measures.

6.2 Influencer Behaviour

This section addresses the first of our core research questions concerning the strategies employed by influencers to attain widespread influence, and the extent to which these strategies are effective or constrained.

Opinion and Out-Strength Evolution

We track the opinions and out-strength of the same set of nodes across both the influencer and non-influencer network scenarios over time. The key difference is that, in the non-influencer scenario, these nodes do not engage in influencer-specific behaviours.

Figure 6.2 shows that influencer agents consistently shift towards extreme opinions, with rapid early movement followed by convergence near ± 4. In contrast, when these agents do not exhibit influencer behaviour, their opinions fluctuate more moderately around ± 1. However, rare instances of extreme deviation occur, with some opinions exceeding ± 4—as illustrated in Figure 6.2b, where one agent's opinion drops sharply to -8.

Figure 6.3 displays the evolution of out-strength among influencer agents, revealing a consistent and substantial increase over time. In contrast, Figure 6.3b shows more erratic trajectories in the non-influencer scenario, with out-strength levels fluctuating and lacking a sustained upward trend.

This disparity is further highlighted in Figure 6.5, which compares the out-strength of influencer agents to the network average. Influencers exhibit out-strength levels approximately 3.4x higher than non-influencer agents, underscoring the effectiveness of their strategy in amplifying influence within the network.

Relationship between Opinion and Out-Strength

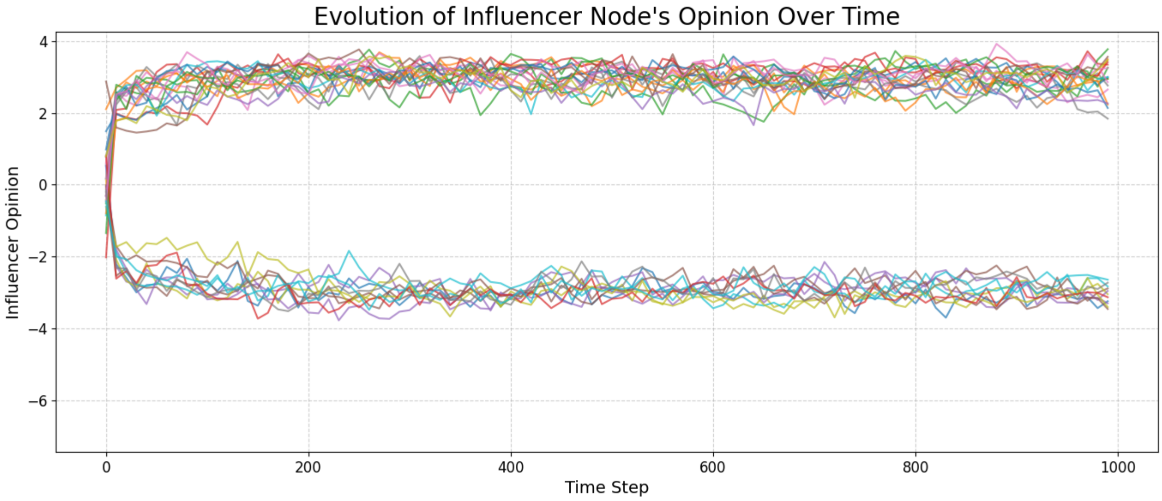
We narrow in on the relationship between opinion and out-strength by examining two specific networks from the non-influencer set in Figures 6.2b and 6.3b, as illustrated in Figure 6.4.

Out-strength increases with opinion extremity in both networks, but begins to decline beyond timestep 600 as opinions become excessively extreme.

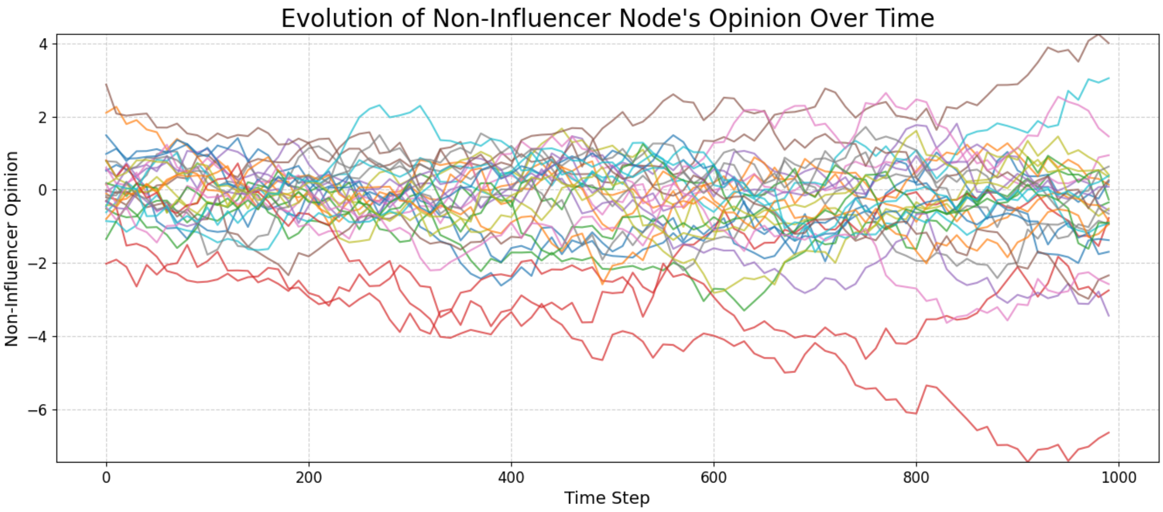
Network 13 In Network 13, the out-strength of the non-influencer node declines from approximately 6500 to 5000 when its opinion falls below -4 around timestep 750. Subsequently, as the node's opinion moderates towards -2, its out-strength rebounds sharply, peaking at around 8000.

Network 23 In Network 23, the agent's out-strength, which had been gradually increasing, begins to decline once its opinion drops below -4 at approximately time-step 600. This downward trend accelerates as the opinion reaches beyond -6 around time-step 800, with the agent's out-strength falling to just above zero.

These findings indicate a non-linear relationship between opinion extremity and network influence. Moderate extremity can enhance an agent's out-strength, but once opinions surpass critical thresholds (e.g., -4 or -5), influence declines sharply.

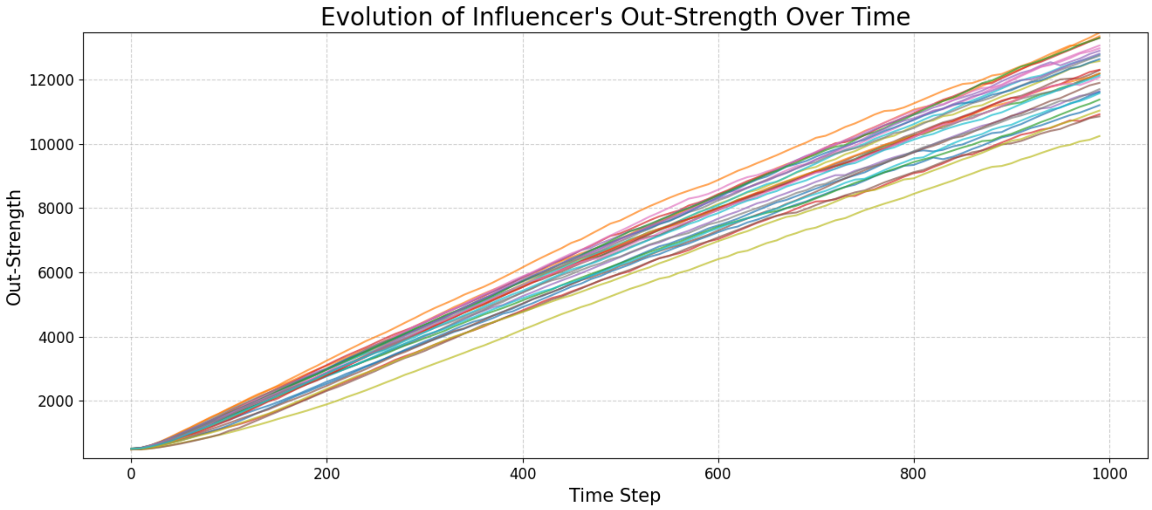


(a) Evolution of opinions over time for influencer nodes across all networks. The results indicate a noticeable shift in opinion towards extremism over time.

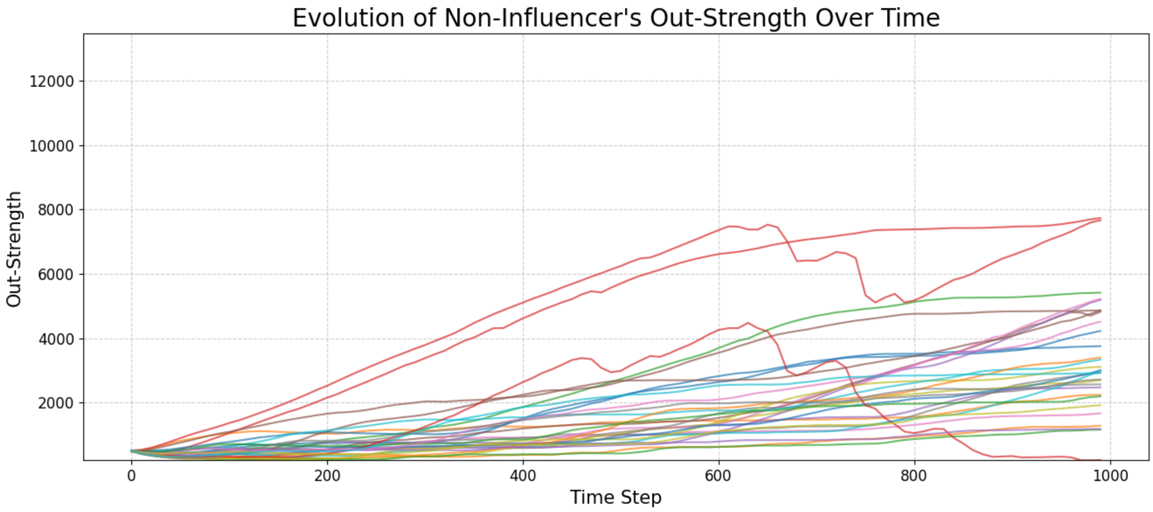


(b) Evolution of opinions over time for the same set of nodes when they do not exert influence beyond their normal interactions. Compared to the previous case, the opinions remain more stable.

Figure 6.2: Comparison of opinion evolution in networks with and without active influencer behaviour. The results illustrate how influencers significantly alter their opinion dynamics compared to when they act only through regular network interactions.



(a) Evolution of out-strength over time for influencer nodes across all networks. Out-strength, representing the total influence exerted, increases significantly over time as influencers strategically shape their opinions.



(b) Evolution of out-strength over time for the same set of nodes when they do not exert influence beyond their normal interactions. The out-strength still increases but at a much slower and weaker rate compared to the previous scenario.

Figure 6.3: Comparison of out-strength evolution in networks with and without active influencer behaviour. The results show that when influencers engage in influencer behaviour, their out-strength grows rapidly, whereas in the absence of deliberate behaviour, this increase is weaker, more gradual and may even decline over time.

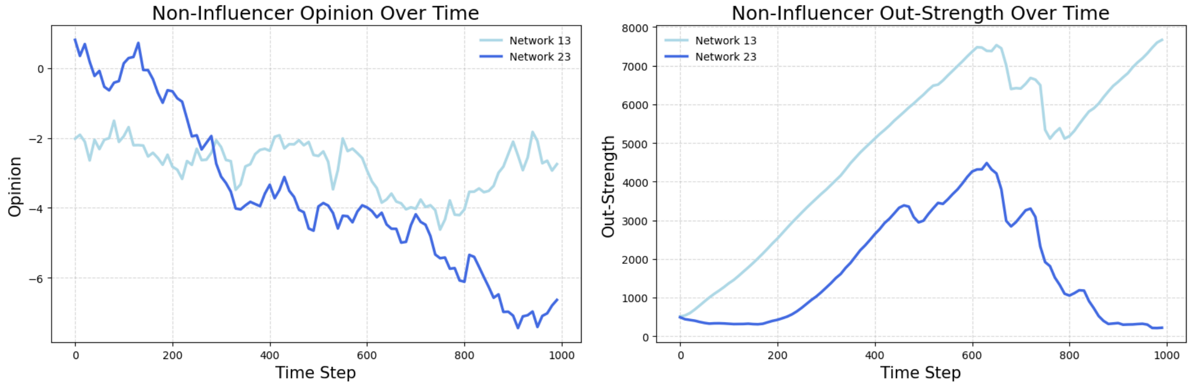


Figure 6.4: Evolution of opinion and out-strength over time for Network 13 (light blue) and Network 23 (dark blue). While out-strength increases with opinion extremity, it declines significantly beyond the threshold of -4.

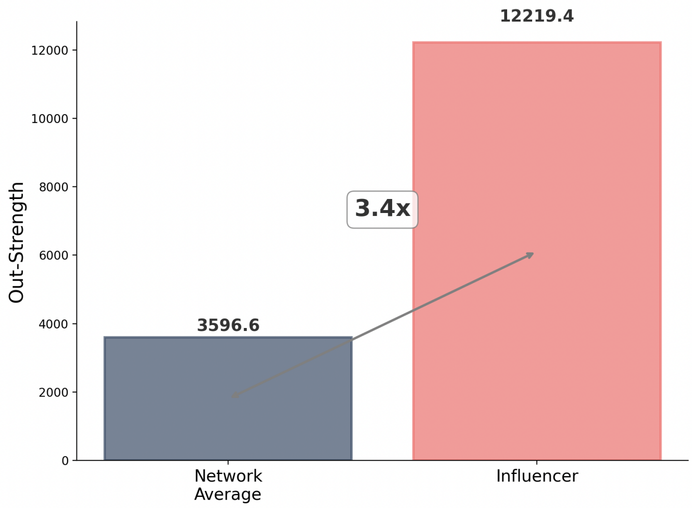


Figure 6.5: Comparison of the average final out-strength between influencers and regular network nodes in networks with influencers. Influencers have 3.4x greater average out-strength than their peers.

6.3 Correlations

In this section, we conduct a correlation analysis to: (1) identify the parameters or variables influencing the behaviour of influencers, and (2) assess whether a single influencer has an impact on the broader network's opinion.

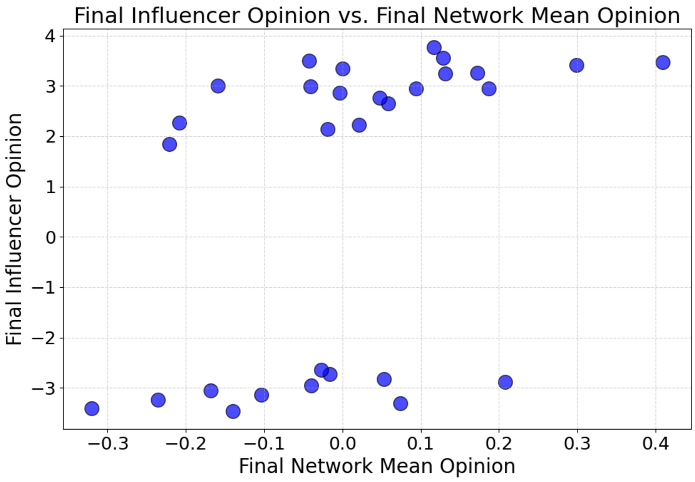
Figure 6.6a shows no significant correlation between the final network opinion and the final influencer opinion, suggesting minimal widespread impact from the influencer.

Figure 6.6b reveals no correlation between the influencer's initial local social norm and their final opinion, indicating that the local social norm does not influence the influencer's final stance.

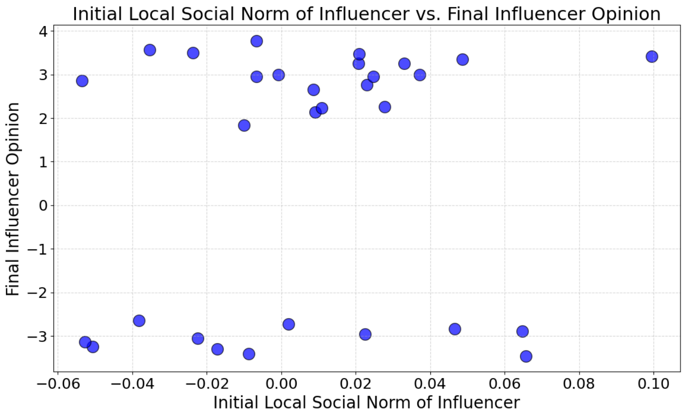
Figure 6.6c further demonstrates no correlation between the influencer's initial and final opinions, suggesting that their final opinion is not influenced by their starting position.

Together, these findings imply that the influencer's final opinion is independent of their initial opinion or local social norm, instead driven by whichever trajectory maximises out-strength.

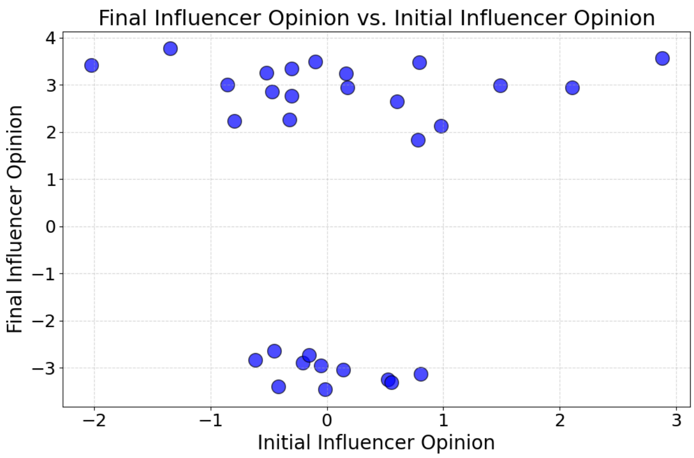
Correlation analysis across all figures found no significant correlations between the outcomes ($p > 0.05$ for all cases).



(a) Scatter plot comparing the final opinions of influencers with the final average opinion of the entire network. Whilst influencers diverge from the mean network opinion, there is no correlation between the network's final opinion and the influencer's.



(b) Scatter plot comparing the final opinions of influencers with their initial local social norms. The results show little correlation.



(c) Scatter plot comparing the final opinions of influencers with their initial opinions. There is no correlation between the polarity of the initial and final opinion.

Figure 6.6: Comparison of correlations between influencer and network opinions and local social norms.

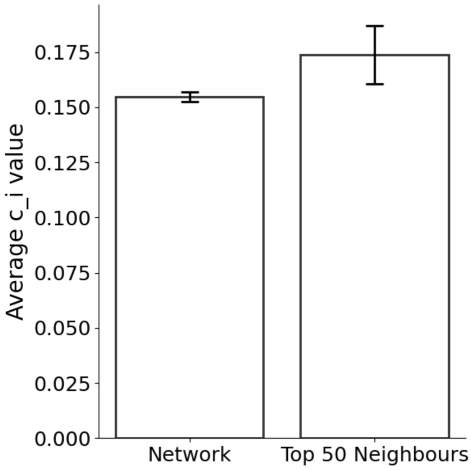
6.4 Top 50 Neighbour Characteristics

We examine the top 50 neighbours of the influencer, the individuals most influenced by the agent.

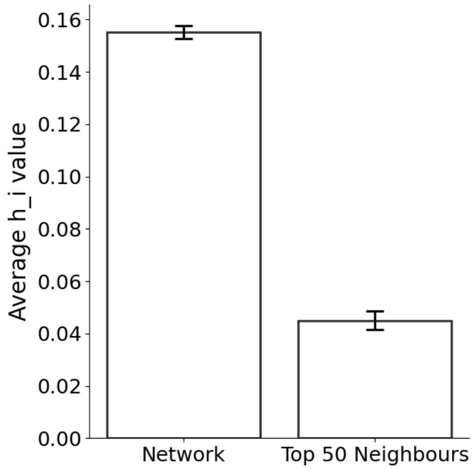
Figure 6.7 compares conformity, homophily, and attention to novelty traits between the overall network and these 50 individuals. Two-sample Welch t-tests revealed statistically significant differences ($p < 1 \times 10^{-19}$) for all traits. Effect sizes, as measured by Cohen's d, indicated stronger effects for homophily ($d = -1.346$) and attention to novelty ($d = 1.346$), with a smaller effect for conformity ($d = 0.229$).

These findings suggest that the most influenced individuals exhibit lower homophily and higher attention to novelty on average, implying that influencers disproportionately affect agents with these traits. The analysis also shows that variation in attention to novelty and homophily has a more substantial impact on influence susceptibility than conformity.

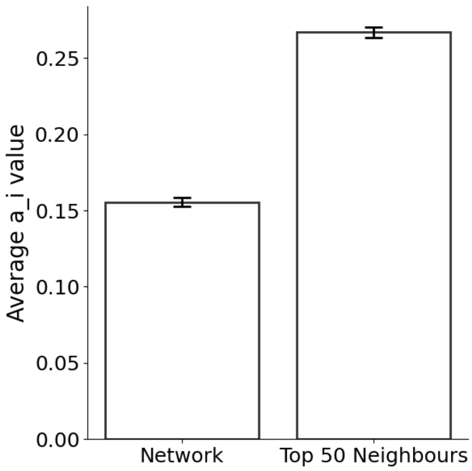
Refer to Table A.3 in Appendix A for a comprehensive summary of the statistical analyses conducted on the c, h, and a values for these nodes.



(a) Average values of parameter c (conformity) for the entire network and for the top 50 neighbours of influencers.



(b) Average values of parameter h (homophily) for the entire network and for the top 50 neighbours of influencers.



(c) Average values of parameter a (attention to novelty) for the entire network and for the top 50 neighbours of influencers.

Figure 6.7: Comparison of conformity, homophily, and attention to novelty parameters for the overall network and the top 50 neighbours of influencers. Highlighting unique social and cognitive characteristics of the most influenced individuals relative to the broader network.

Figure 6.8 plots the network nodes based on their individual levels of homophily and attention to novelty, with each node coloured according to the level of influence exerted by the influencer. The pattern confirms that individuals with lower homophily and higher attention to novelty are more susceptible to influence, reinforcing that the influencer's effect is concentrated among agents with these traits.

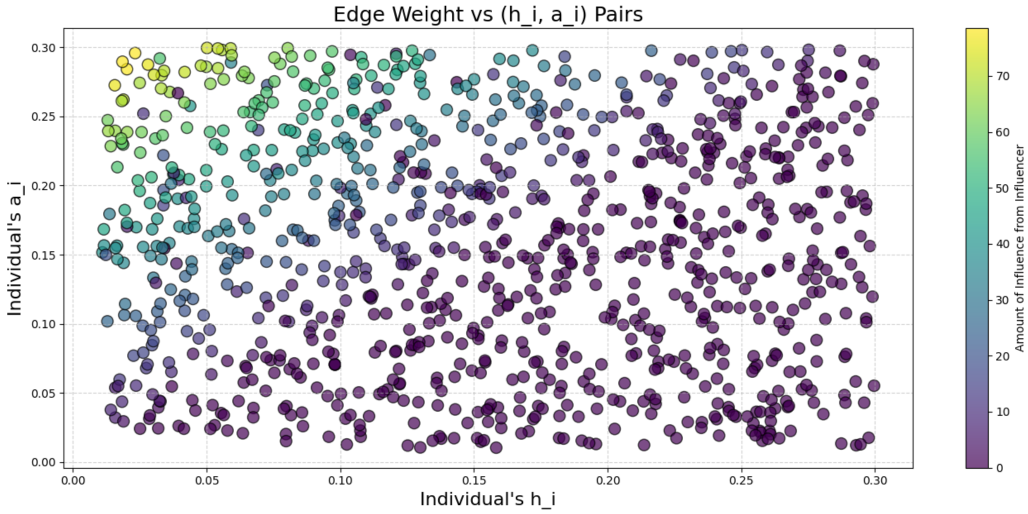


Figure 6.8: Plot of an influencer's strength across different parameter value pairs, where each dot represents a unique individual indexed by its h_i, a_i value. The x-axis corresponds to the homophily parameter (h) and the y-axis represents the attention to novelty parameter (a). The colour of each dot corresponds to the strength of influence from the influencer, with darker or lighter shades representing varying levels of influence. The influencer's strength varies with these edge characteristics and is greatest for low homophily-high attention to novelty pairs.

6.5 Top 50 Neighbour Opinions

We examine the opinion distributions of the top 50 most influenced neighbours, both with and without the presence of an influencer. Figures 6.9, 6.10, and 6.11 use Kernel Density Estimators (KDE) plots, which smooth discrete histograms to enable more accurate comparisons with large population-level distributions.

Figure 6.9 shows the distribution of absolute opinion values for these neighbours, contrasting scenarios with (red plot) and without (blue plot) an influencer. In the absence of an influencer, the distribution closely follows a normal curve around zero, reflecting a concentration of moderate opinions (or consensus).

With the influencer present, the distribution becomes more bimodal. A smaller peak aligns with the moderate cluster seen in the blue distribution, while a larger peak emerges around an opinion value of 3. This shift indicates that the influencer's presence drives a greater concentration of opinions towards more extreme values.

Additionally, the red curve remains elevated relative to the blue curve in the high-opinion range (beyond 5), particularly near the upper limit of 8. This suggests that the influencer not only shifts opinions towards greater extremity but also broadens the range of conviction among the most influenced agents. We will explain this effect in more detail in Chapter 7.

Figure 6.10 illustrates the effect of influencer presence on the extremity of opinions among the top 50 most influenced individuals, measured as the absolute difference between an individual's opinion and the network average. In the non-influencer scenario, most opinions cluster within 0-1 points of the average, reflecting moderate consensus. In contrast, with influencers present, the distribution shifts markedly, showing a concentration of opinions around 3 points away from the average, indicating a substantial increase in extremity. This difference is statistically significant ($p = 1 \times 10^{-42}$), with the average opinion distance rising from 1.13 to 3.02 in the presence of influencers.

Figure 6.11 compares the opinion distances between network nodes and the influencer's final opinion, both with and without active influence. This shows us how close the nodes would actually be to the influencers extreme opinion if there were no influence effects (blue plot). In the absence of active influence, most individuals maintain opinions 2–4 units away from the influencer's stance, reflecting moderate divergence. When the influencer is active (red), the distribution shifts notably towards smaller distances, with a larger proportion of individuals aligning with the influencer's more extreme views. This shift confirms that the movement towards extremism is a direct result of the influencer's behaviour. A statistically significant difference is observed ($p = 6.2 \times 10^{-95}$), with the average opinion distance decreasing from 3.24 in the inactive condition to 1.76 when the influencer is actively exerting influence.

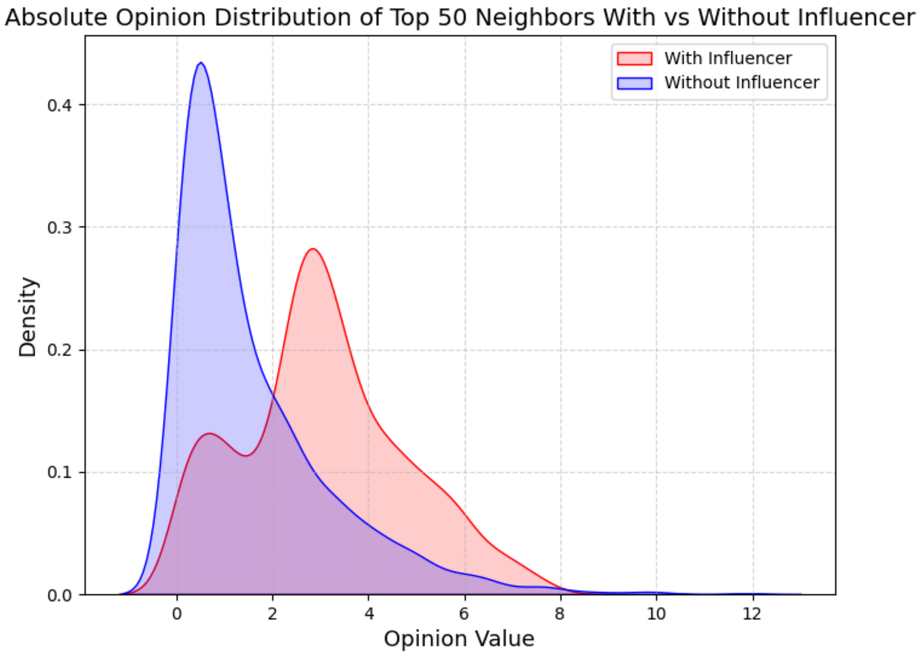


Figure 6.9: Absolute opinion distributions of the top 50 neighbours, comparing scenarios where the influencer is either actively shaping opinions or inactive.

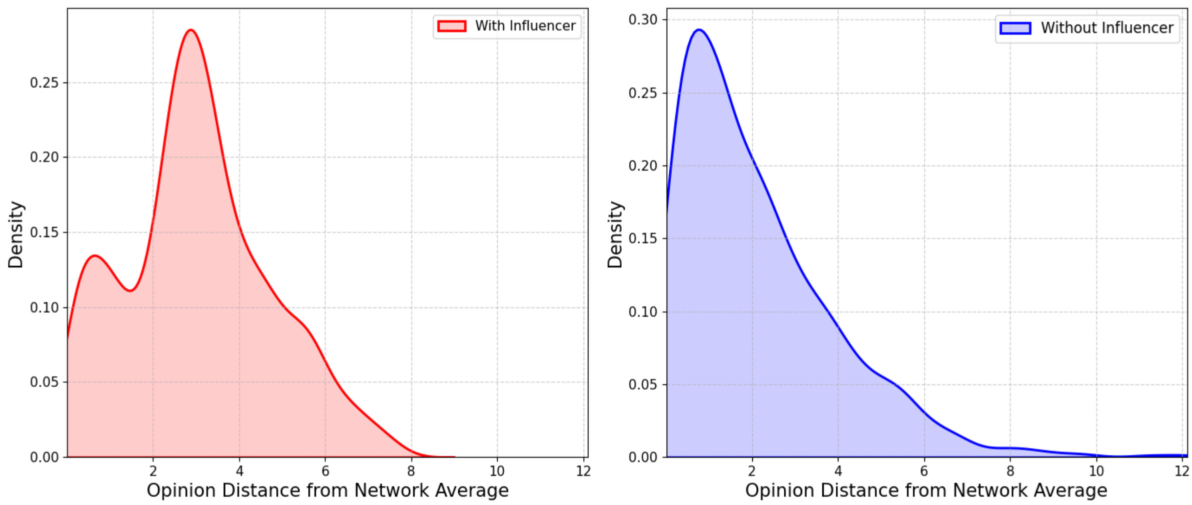


Figure 6.10: Opinion distances of the top 50 neighbours from the network average in scenarios where the influencer is either actively shaping opinions or inactive. The red plot represents networks with active influencer behaviour, while the blue plot shows networks without it.

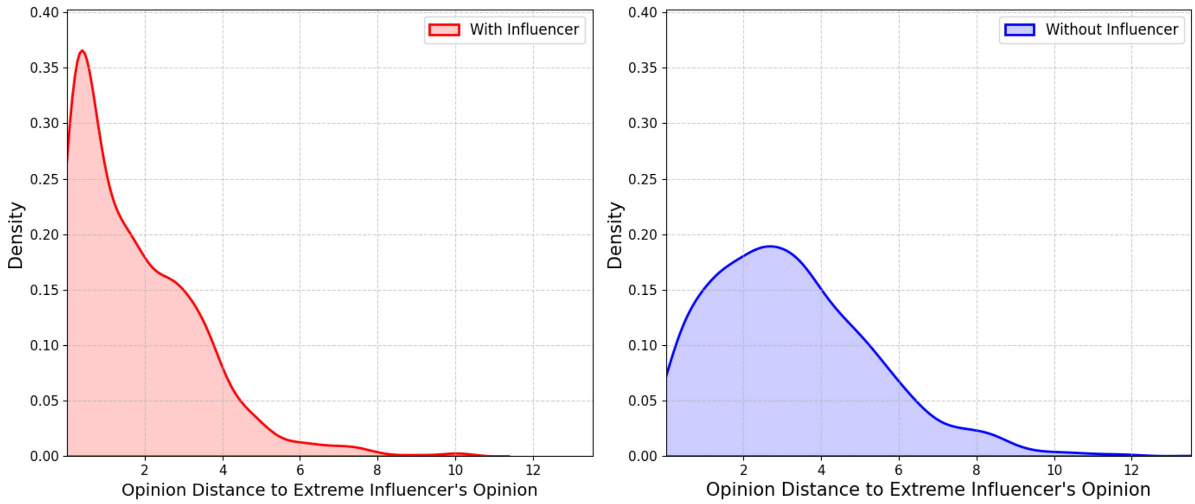


Figure 6.11: Opinion distances of the top 50 neighbours from the influencers opinion, comparing scenarios where the influencer is either actively shaping opinions or inactive. The red plot represents networks with active influencer behaviour, while the blue plot corresponds to networks without it.

6.6 Multiple Influencers

Figure 6.12 presents a comparison of the final opinion distributions for networks with and without 100 influencers. In the absence of influencers, the opinions are primarily centred around the neutral value of 0, with a smooth, bell-shaped distribution that tapers off as opinions move towards more extreme values. This indicates consensus within the population.

However, in the network where 100 influencers are present, a notable shift occurs. While a small fraction of the population remains centred near 0, the distribution becomes more fragmented and exhibits distinct peaks at absolute opinion values around 2 and 4–5. This shift suggests that the influencers are driving the population towards more polarised and extreme opinions, resulting in a network that is increasingly divided between moderate and highly extreme positions. Statistically significant differences were observed between the two conditions ($p = 4.3 \times 10^{-6}$), with the influencer network exhibiting a markedly higher prevalence of extreme opinions.

Figure 6.13 shows the distribution of opinions across the networks. Opinion categories (Strongly Negative, Moderately Negative, Neutral, Moderately Positive, Strongly Positive) are defined relative to the mean and standard deviation of opinions: Strongly Negative opinions are more than 1.5 standard deviations below the mean, Moderately Negative between 1.5 and 0.5 standard deviations below, Neutral within 0.5 standard deviations, Moderately Positive between 0.5 and 1.5 above, and Strongly Positive more than 1.5 above the mean. The non-influencer network serves as the baseline for comparison against the more polarised influencer network. It's important to note that the observed asymmetry between strongly positive and strongly negative individuals is not inherent to the model, but rather a consequence of a higher proportion of influencers in this one scenario adopting negative viewpoints.

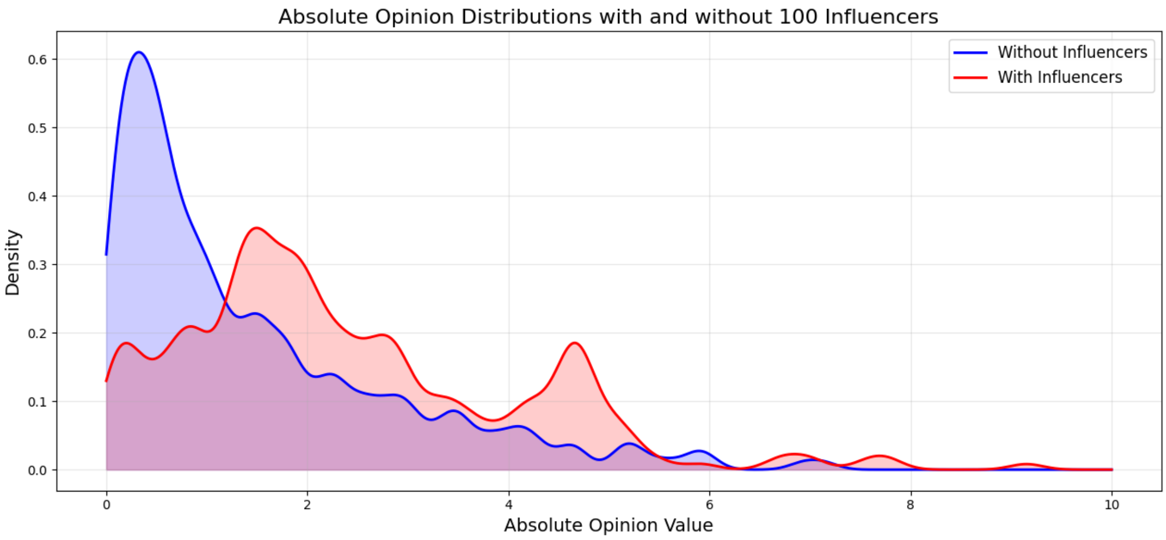


Figure 6.12: Absolute opinion distributions in networks with and without 100 influencers.

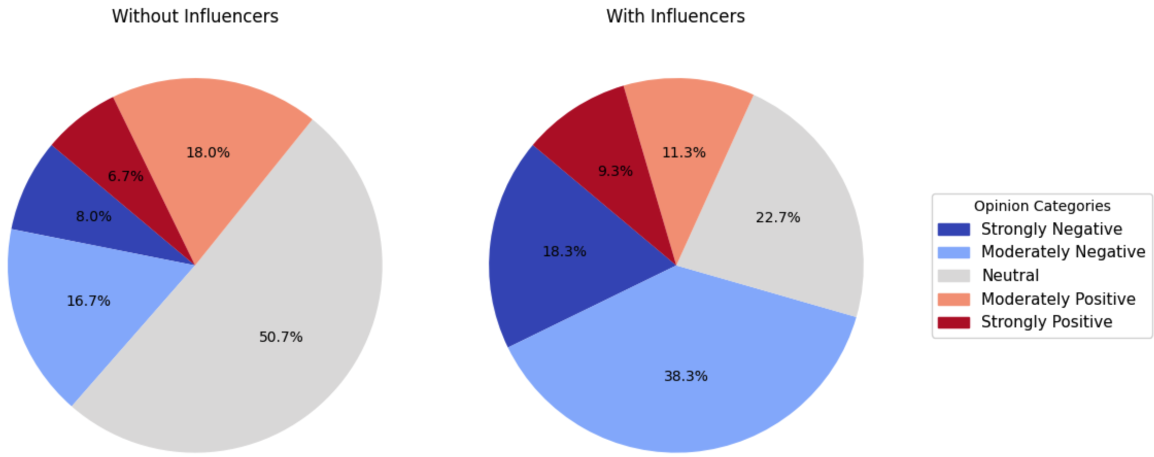


Figure 6.13: Distribution of agent opinions under two conditions: with and without the presence of 100 influencers. Categories are defined relative to the mean and standard deviation of the opinion distribution from the non-influencer network. The inclusion of influencers shifts the opinion landscape, increasing polarisation as evidenced by the rise in extreme (strongly negative and strongly positive) positions.

Chapter 7

Discussion

This section discusses the key findings from our results. Influencer opinions quickly become more extreme with a clear positive correlation, up to a point, between opinion extremity and out-strength. Influencers most significantly affect individuals with low homophily and high attention to novelty, pushing their neighbours' opinions towards extremity. Networks with multiple influencers exhibit higher average extremity and greater polarisation compared to those without.

We aim to answer our key question by exploring the implications of these findings for opinion dynamics: to what extent can they shed light on the drivers, mechanisms, and consequences of extremism and polarisation linked to online influencers?

7.1 Influencer Behaviour

This section will address key question (1): what strategies do influencers employ to gain widespread influence, and how effective are these?

7.1.1 Influencers Favour Extremism

Our model shows that influencers tend to gravitate towards extreme positions as a strategy to maximise influence. This shift towards extremism stems from the balance between social acceptance and the need for novelty. Influencers increase their impact by adopting opinions that diverge significantly from the average, since this distinction draws more attention. However, this strategy comes with limitations; too much extremity risks severing ties with a large portion of their network. The resulting opinion extremes generally stabilise within a narrow range, reflecting the tension between amplifying influence and maintaining social connections.

We now address several key questions: 1) Why do influencers adopt this particular strategy? 2) What accounts for the continuous linearity of an influencer's out-strength, even after reaching the plateau? 3) Why does the opinion limit fall between 4 and -4, rather than within the intuitive range of -5 to 5, as suggested by the attention limit constraint?

Why the Influencer Chooses the Extremist Pathway

Our influencer optimally maximizes its influence across all neighbours by maximising, where possible, the value of $w_{i,\text{inf}}$ for each neighbour i, as defined in Eq. 2.2.6.

Each value of $w_{i,\text{inf}}$ depends on both F_h and F_a. To increase $w_{i,\text{inf}}$, both F_h and F_a need to be positive, which requires that the distance $|x_i - x_{\text{inf}}| < \theta_h$ and $|\langle x \rangle_i - x_{\text{inf}}| > \theta_a$. The influencer has control over only the value of x_{inf}, which represents the influencer's own opinion.

Given the equation for F_h, the maximum value of F_h occurs when $x_i = x_{\text{inf}}$, resulting in $F_h = \theta_h$. Thus, the potential size of F_h is limited by θ_h, and consequently, the potential increase in w_{ij} is also constrained.

While the influencer could adopt the average opinion $\langle x \rangle_i$ to maximise F_h by being as similar as possible to as many individuals as possible, this is unlikely to be an optimal technique. Since each subsequent increase in influence is limited by the threshold, the potential for increasing inter-personal influence through homophily is not optimal.

On the other hand, for F_a to be positive, the value of $|\langle x \rangle_i - x_{\text{inf}}|$ must be greater than θ_a. Since there is no upper limit on the size of x_{inf}, the value of $|\langle x \rangle_i - x_{\text{inf}}|$ also has no upper bound as long as it exceeds θ_a. Therefore, F_a is unlimited in its potential size, and consequently, the potential increase in $w_{i,\text{inf}}$ is unlimited.

Therefore, it makes strategic sense for the influencer to prioritise novelty when selecting the value of x_{inf}, as the maximum potential increase in $w_{i,\text{inf}}$ for their neighbours occurs when F_a (attention to novelty) is maximised, rather than F_h (homophily). Since $|\langle x \rangle_i - x_{\text{inf}}|$ is maximised when x_{inf} is as different as possible from $\langle x \rangle_i$, any large value of x_{inf} will be effective. This explains why the influencer may adopt extreme opinions. By diverging significantly from the average opinion, they cumulatively enhance their influence over others.

However, the constraint arising from the relationship

$$w_{i,\text{inf}} = 0 \quad \text{when} \quad |x_i - x_{\text{inf}}| \geq 5$$

introduces an important consideration for the influencer's strategy. If the influencer's opinion is too far from an individual's opinion then$w_{i,\text{inf}} = 0$, and no influence is exerted over that individual. The influencer must select a value of x_{inf} that is sufficiently distinct to maximise the novelty factor, but not so extreme as to risk losing ties with others holding differing opinions. Succinctly, there is a critical balance between social tolerance and divergence here. This summarises why the influencer, mathematically, decides to take the extremist route.

Why the Influencer Continues to Grow after Plateauing its Opinion

When the influencer's opinion plateaus at an extreme value of x_{inf}, their out-strength continues to increase because their opinion remains distinct from others in the network. Specifically, as long as the difference $|\langle x \rangle_i - x_{\text{inf}}|$ exceeds the threshold θ_a, the influencer's opinion remains unique enough to exert an increasing amount of influence.

For agents whose local social norms are not dominated by the influencer's extreme opinion, x_{inf} remains distinct enough to continue increasing the influencer's influence, even when it plateaus (e.g., at 4 or -4). Given the largely centrist population, most agents maintain local social norms that differ significantly from the influencer's extreme views, which further amplifies their influence.

Why the Limit is [4,-4]

As previously noted, the opinion range is constrained to $[-4, 4]$, rather than extending to $[-5, 5]$.

To illustrate this limitation, we analyse the final opinion distribution across the population. On average, across 30 simulations, 75.2% of nodes exhibited opinions greater than -1, and 50.5% held opinions greater than 0 at the end of the simulation. This outcome is a consequence of the initial normal distribution, along with the conformist behaviours that ultimately drive consensus within the system around the central point.

If the influencer were to extend their influence to an opinion of 5 instead of 4, they would neglect the 25% of individuals whose opinions fall within the range of $[-1, 0]$. This results from the distribution of opinions in our network, where there is a higher concentration of nodes within the $[-1, 1]$ range compared to those outside it. Also, due to the attention limit constraint, an opinion of 5 effectively cuts off relationships with individuals whose opinions are less than 0. As a result, it is more logical to set the maximum opinion value at 4 and the minimum at -4, as this maximises influence over a larger proportion of the population.

7.2 Vulnerable Agents

This section addresses key question (3): is the influencer's impact broad and diffuse, or more focused and targeted?

7.2.1 Vulnerable Agents Share Unique Characteristics

As shown in Figure 6.7, the most influenced agents typically exhibit **higher attention to novelty and lower homophily**. In other words, influencers are most effective at capturing the attention of individuals who are particularly drawn to radical, extremist nodes, and who are more receptive to differing opinions. This combination of traits creates an ideal environment for influencers to exploit extremism as a means of expanding their influence.

Homophily

In both Figure 6.7b and Figure 6.8, we identified the most susceptible individuals as those with significantly lower homophily (h_i) levels compared to the rest of the population. This suggests low homophily is a key indicator of an easily influenced individual. Since influencers tend to adopt extreme opinions, the value of F_h in Eq. 3.4 would be large and negative, akin to homophily pushing dissimilar opinions away. However, a lower h_i value weakens the effect of F_h, reducing its effect on $w_{i,\text{inf}}$. This enables influencers to maintain strong connections with these individuals, despite holding extreme opinions.

Attention to Novelty

As with homophily, we identified the most susceptible agents as those with significantly higher attention to novelty (a_i) levels. As shown in Eq. 3.5, the difference between $\langle x_i \rangle$ and x_{inf} is likely to surpass the threshold given the large value of extreme opinions adopted by our influencers. This results in a large, positive F_a value. A higher a_i value amplifies the effect of this difference on $w_{i,\text{inf}}$, increasing the influence that an extreme node, such as our influencer, exerts over node i.

Conformity

The influencer focuses on relationships (and thus edge weights) where conformity does not play a determining role in the value of influence between two nodes. This is primarily driven by the other parameters, homophily and attention to novelty. As a result, there is no correlation between conformity and the strength of the influencer.

The Spread of Extremism in Vulnerable Agents

Our results show that the presence of an influencer significantly increases the likelihood that vulnerable agents adopt more extreme opinions. These agents' traits make them especially open to influence, giving the influencer disproportionate power over them. In a conformist system, this tilts their perceived social norm towards extremism, suppressing the effects of more moderate agents. As this skewed norm becomes the reference point, agents shift their opinions accordingly.

This is captured in the opinion update equation (Eq. 3.8), where an agent's opinion x_i moves toward the local norm $\langle x \rangle_i$. When an influencer dominates the neighbourhood N_i, their extreme stance distorts the norm, and conformity (c) drives x_i in that direction—amplifying the agent's drift to extremism.

Formally:

$$\langle x \rangle_i = \frac{\sum_{j \in N_i} w_{ij} x_j}{\sum_{j \in N_i} w_{ij}}$$

When an influencer dominates the neighbourhood, i.e., $w_{i,\text{inf}}$ is sufficiently larger than all other w_{ij}, we have:

$$\langle x \rangle_i \approx x_{\text{inf}}.$$

Thus, we can rewrite the opinion update equation (Eq. 3.8) as:

$$\frac{dx_i}{dt} = c\,(x_{\text{inf}} - x_i) \tag{7.1}$$

Over time, some nodes continue to grow in extremity, sometimes surpassing the influencer. This is driven by a feedback loop: as the social norm shifts further towards the extreme, more radical opinions become interesting, accelerating the move towards extremism. Additionally, the attention limit filters out moderate viewpoints, while increasing exposure to more extreme opinions on the same side of the spectrum.

Since each node only considers opinions within a fixed 5-point range (as defined in Eq. 4.1), its exposure to alternative views becomes increasingly limited as it drifts towards one extreme. The closer it gets to the ideological edge, the more alternative opinions fall outside its attention range, creating a manufactured asymmetry in exposure. This self-reinforcing process traps the agent in a narrowing opinion exposure channel, amplifying its radicalisation.

Evidence suggests that exposure to extremist content can heighten individual polarisation, making our model a plausible reflection of real-world dynamics (Borchers, 2025). However, the causal link between influencer extremism and audience attitude shifts remains unclear and is still under active investigation (Gentzkow and Shapiro, 2018; Glazer and Wells, 2019; Muñoz et al., 2024).

7.3 Network Impact

Finally, this addresses both key questions (2) and (3): how does the network respond to influencers and the is the influencers impact broad or focused?

Influencer Impact is Limited by their Own Extremity

Network traits remained largely unchanged with or without influencers present. Key metrics from Bullock and Sayama (2023) showed no statistical differences, and our correlation analysis found no significant link between final network opinion and final influencer opinion. This suggests that a single influencer has limited capacity to shape overall public opinion.

The reason lies in how individuals respond to extremism. Moderately homophilic agents with low attention to novelty are less susceptible to influence, especially from extreme views. These nodes deprioritise connections that deviate too far from their social norms. Due to attention limits, agents with opposing opinions often ignore the influencer entirely.

This reveals a trade-off: the more extreme the influencer, the narrower their reach. But within a vulnerable subgroup, agents highly susceptible to influence, the influencer has outsized impact. These relationships are highly imbalanced, often pulling vulnerable agents towards the influencer's position. So while broad influence is limited, concentrated extremism emerges.

Later analysis shows that as the number of influencers increases, the network becomes progressively more polarised. So while extremity limits individual reach, collective extremism among multiple influencers reshapes the network.

Influencer Impact is Limited by the Underlying Population Traits

High homophily individuals tend to form strong bonds with like-minded peers while distancing themselves from more radical opinions. This selective engagement limits the influencer's reach, as their extreme ideas fail to resonate with closed, similarity-driven individuals. Low attention to novelty further diminishes the influencer's effectiveness. Agents indifferent to new ideas or perspectives are unlikely to engage more with extreme or unfamiliar positions. Consequently, in a network whose population is characterised by higher homophily and lower attention to novelty, the influencer's ability to sway that population would be significantly reduced (see Figure 6.8).

Clearly then, the distribution of agent traits within a network sets a limit on the influencer's potential success. The importance of population traits in shaping influencer effectiveness is consistent with findings in other studies like D. Watts and Dodds (2007) and Brooks and Porter (2020).

We can draw a connection between our findings and a social experiment by Homans (1958), which categorised individuals into high and low attraction groups based on how well they got along. A paid participant adopted an opposing opinion to the group average, and after reassessment, 20% of individuals in low-attraction groups shifted towards this opinion, compared to just 7% in high-attraction groups.

If we consider what it means to be less socially bonded in our model, this could be synonymous with agents who, due to low homophily, are more easily pulled away from the social norm, much like the vulnerable agents in our model. This suggests that less socially bonded groups, akin to vulnerable agents, are more susceptible to external influence. These results align with our model, suggesting influencer effectiveness is constrained by the people it interacts with.

We discuss in the next sections the potential this idea offers for reducing the prevalence of polarisation and extremism in society.

7.3.1 Multiple Influencers

Introducing multiple influencers led to large-scale polarisation, replacing our consensus-driven, centrist society with a divided one. This aligns with findings from Brooks and Porter (2020), which suggest that while a single influencer has a focused impact, multiple influencers can cause profound shifts. Similarly, Coculescu et al. (2023) found that competing influencers (from opposing ends of the spectrum) significantly increased polarisation.

In a consensus-driven society, opinions are normally distributed around the centre, with only a few individuals holding extreme views. This reflects a stable societal consensus. However, polarisation divides opinions toward both extremes, widening the gap between them and making it harder to find common ground (Axelrod et al., 2021; Bail, Argyle, Brown, Bumpus, et al., 2018; McCoy, 2019).

Our model reflects this: without influencers, the network gravitates towards neutrality, with 50.67% of nodes remaining neutral. With multiple influencers, neutrality drops to 22.67%, and extreme views double, from 14.67% to 27.66%, signalling a hardening of opinions (numbers are derived from Figure 6.13).

The network structure also evolves, with assortativity rising from 0.15 to 0.30 (see Appendix A - Figure A.2). This increase suggests individuals are more likely to connect with others who share similar opinions, reinforcing stronger intra-group ties. When extreme opinions dominate, these connections further skew social norms, driving opinions further from the mainstream. This dynamic mirrors the creation of echo chambers (self-reinforcing clusters of like-minded individuals) which have been widely studied in online contexts (Baumann, Lorenz-Spreen, et al., 2020; Vicario et al., 2016).

The shift toward extremism is directly driven by the strategies employed by the influencers. Notably, influencers stabilise around $[6, -6]$ instead of $[4, -4]$ (shown in Appendix A - Figure A.1). It's clear that the reach of the influencers no longer aims to encompass the middle, moderate section, which is now more polarised than before. Instead, influencers focus on propagating more extreme opinions, targeting their peers on their side of the spectrum to further radicalise them. This shift highlights the self-reinforcing nature of polarisation, where the process not only deepens but also accelerates. Similar dynamics are observed in real-world contexts, where extremism continually feeds back into itself (Bail, Argyle, Brown, and Volfovsky, 2018; McCoy, 2019; Panizza et al., 2021).

7.4 Real-World Consequences and Ideas

Our model gives us a pathway in understanding how political polarisation may come about through multiple influencers. As a single influencer becomes more extreme, their effect on the majority weakens, but their influence on vulnerable agents intensifies, deepening polarisation within smaller groups. In contrast, multiple influencers drive widespread polarisation across the network.

This section highlights comparisons for an interdisciplinary discussion of the model's findings, seeking to address ideas from key question (4).

Radicalised Subgroups and Their Risks

Single influencers in our model often radicalise a small subset of individuals, a dynamic with serious real-world implications. Historically, small, extreme groups have driven violence, terrorism, and unrest (Centola et al., 2018; Osnos, 2019). Thus, even narrower influence can be dangerous when it fosters deep ideological commitment. This underscores the risk posed not by widespread persuasion, but by the creation of smaller, radicalised pockets. Our model may then act as a push towards the identification of such groups online, recognising their potential impact.

Psychological Profiles of At-Risk Individuals

In our model, the most susceptible individuals, those with low homophily and high novelty-seeking, closely resemble real-world profiles of individuals vulnerable to radicalisation. The identification of at-risk agents based on psychological traits aligns with strategies used by online policing and charity organisations. This represents a unique contribution, as existing opinion dynamics models rarely emphasise how individual characteristics critically shape broader network effects.

Research links social isolation and weak social ties to vulnerability to extremist ideologies (Horgan, 2017; McCauley and Moskalenko, 2008; Simi and Futrell, 2010). Low homophily reflects this isolation, making individuals more open to radical influence due to a lack of stabilising group norms.

Recognising these traits can inform targeted interventions aimed at preventing online radicalisation and reducing polarisation, although this would require greater empirical validation.

Critical Mass and the Collapse of Consensus

Our model shows that when multiple influencers adopt extreme views, network-wide polarisation emerges. A critical mass of extreme, influential agents fractures the network into opposing clusters, making consensus harder to achieve.

This aligns with findings from other agent-based models (Betts and Bliuc, 2022; Ding et al., 2023) and U.S. election Twitter/X analysis, which links influencer extremity to heightened user polarisation (Flamino et al., 2023). However, the mechanisms through which multiple influencers drive polarisation are poorly understood (Betts and Bliuc, 2022). While research often focuses on extremist content or the growing tendency of influencers to promote it, few empirical studies directly link the influence of multiple extreme influencers to large-scale societal polarisation (Betts and Bliuc, 2022; Hassan et al., 2018; Nöfer, 2024; Schleffer and Miller, 2021).

Further research using large-scale social media datasets could clarify how multiple influencers contribute to societal division as indicated by our model.

7.4.1 Insights and Transferable Ideas

Building on our model and its alignment with real-world dynamics, we outline actionable strategies to counter online radicalisation and polarisation.

Identifying Vulnerable Individuals

Certain traits make individuals more susceptible to extremist influence. Our model highlights two key risk factors:

- **Low Homophily:** Individuals experiencing social isolation, loneliness, or withdrawal often seek connection in online spaces, making them more open to radical ideologies.
- **High Novelty-Seeking:** Impulsive or attention-seeking behaviours correlate with higher engagement with extreme or conspiratorial content, increasing exposure to radical views.

Recognising these traits can guide early interventions and moderation strategies online.

Platform-Level Moderation Strategies

To limit the spread of polarisation, social media platforms should adopt proactive moderation approaches:

- **Track Extremity Trends:** Monitor influencers for shifts towards extreme content. Escalating rhetoric can signal growing community radicalisation and potential risk.
- **Promote Bridging Content:** Algorithms should amplify content that connects opposing viewpoints and promotes dialogue. This helps counterbalance ideological silos and reduce polarisation as influencer presence scales.

These strategies, rooted in our model, offer a foundation for mitigating the real-world impact of online extremism.

Chapter 8

Critical Analysis

In this chapter, we take a step back from the previous discussion of real-world dynamics and models to critically assess the underlying assumptions and limitations of our framework. While our model offers valuable insights into the influence of agents and the process of polarisation, it simplifies certain aspects that could be expanded to better reflect the complexities of real-world influence.

8.1 Research Approach

Our initial analysis focused on a single influencer within a network of $n = 1000$ agents, providing valuable micro-level insights into influencer behaviour. However, this limited the broader applicability of our findings. Real-world social media ecosystems typically involve multiple influencers, whose inclusion reveals more complex dynamics, including overlapping influence and widespread polarisation. A more comprehensive approach, incorporating multiple influencers from the outset, would better capture these phenomena. Future work could explore variations in influencer interaction, influence overlap, and the resulting network-wide polarisation under different configurations of multiple influencers, followed by more extensive statistical analysis we afforded to single influencers.

8.2 Weaknesses of the Model

Simplistic Opinions

Our current model gives each node i just one opinion value x_i. Homophily and attention to novelty only look at differences in this value meaning the influence relationships between individuals are based on a single dimension of their views. Realistically, the dynamics of influence will rely on a variety of opinions on different topics. Therefore, a better model may consider a node i's set of opinions $\mathbf{x}_i = (x_1, x_2, x_3, \ldots, x_n)$, where each x_j represents the opinion of node i on a different topic. This reflects a broad spectrum of opinions: pasta preferences, climate change views, or their enjoyment of computer science. This adds a layer of realism to how inter-personal influence relationships develop. While you might share a similar pasta preference with someone, it doesn't necessarily mean you will be equally influenced by their views on immigration or climate change. In contrast, our current model assumes that a single opinion defines the entirety of how influence is exerted between individuals, which oversimplifies the complexity of real-world interpersonal influence.

Hard Thresholds

The thresholds used in our edge update equation, see Eq. 3.7, suggests humans accept or reject opinions using a hard threshold function. Tolerance for extreme opinions is likely to vary significantly between individuals. Some people may be more willing to engage with extreme ideas, while others may quickly dismiss them. This means openness should be defined heterogeneously just like our h_i and a_i values. By introducing heterogeneous thresholds for attention to novelty and homophily (θ_a and θ_h), we may better model the complexities of cognitive openness.

The same can be said for our attention limit which cuts off relationships that differ in a '5-point' range. In reality, this tolerance is likely to vary across individuals. However, this simplification may still reflect a key feature of the digital era. Algorithmic content systems often filter out content that diverges too far from a user's established views. In effect, these systems may enforce a form of attention limit on behalf of the user, suppressing exposure to more extreme or differing opinions. Thus, while the uniform cut off may seem overly rigid, it may realistically approximate the influence of algorithmic gatekeeping in online environments.

Importance of Topics

Individuals may assign varying degrees of significance to different topics. In our model, x_i is treated uniformly across all topics, but in reality, individuals may be strongly influenced by opinions on one subject (e.g., climate change) while being less affected or even resistant to views on another (e.g., dietary preferences). Introducing topic-dependent influence weights would more accurately capture the selective nature of belief adoption and the varying susceptibility to influence across different opinion areas.

Personality Variation

Furthermore, personality traits play a significant role in shaping how relationships form. For example, homophily influences relationships not only based on shared opinions but also on more enduring characteristics such as an individual's background, age, race, and gender (McPherson et al., 2001). In contrast, our model only accounts for variation between agents in terms of attention to novelty and homophily, traits that are focused solely on how individuals process and adopt opinions, and do not reflect broader aspects of identity or social characteristics. This limitation restricts the model's ability to fully capture the complexity of how relationships form, even in online spaces, where these broader traits also influence interactions.

Exogeneous Influencers

Finally, in extending our model to include influencers, we introduced an exogenous mechanism that disrupts the main opinion dynamics simulation. Each time the influencer acts, the simulation pauses to allow the influencer to assess the network and generate content. This halt increases computational time and introduces a conceptual inconsistency. In reality, influencers cannot pause time; while they are preparing content or formulating a strategy, the rest of the network continues to evolve. Opinions shift, relationships change, and dynamics unfold independently. Freezing the system during influencer updates oversimplifies this complexity and risks misrepresenting the temporal interplay between influencers and the broader network. Future work should more realistically model influencer activity as asynchronous and continuous, and compare runtime costs under both synchronous and asynchronous update schemes.

A Restricted World

Our simulation uses a static network of 1000 fully connected nodes, with no entry or exit over time. While this simplifies implementation, it fails to capture the fluidity of real-world social systems, where individuals regularly join, leave, or disengage. The absence of demographic turnover or network churn limits the generalisability of our results, particularly for population-level dynamics.

Dynamic churn can be partially modelled within a fixed population by resetting node weights and randomising opinions, approximating entry and exit without altering network size. However, scaling beyond a fixed node count introduces more fundamental issues. Fully connected networks incur a computational complexity of $O(n^2)$ per timestep, with runtime increasing as n grows. This is exacerbated in multi-influencer models, where the addition of exogenous influencers increases influence pathways and computational load. For instance, simulating a network with $n = 300$ and 30 influencers required 4.4 hours of computational time.

These constraints underscore a key trade-off in large-scale simulation: balancing computational feasibility with representational realism.

8.3 Weaknesses of the Influencers

Limitations of the Influencer Model

A key limitation of our framework lies in the computational implementation of influencer strategy. Influencers are restricted to a greedy optimisation process aimed at maximising immediate out-strength, which inherently favours attention-seeking through novelty. This dynamic progressively drives influencers toward extreme positions.

Crucially, this tendency toward extremism is not an abstract artifact, but a direct outcome of the interaction mechanisms embedded in the model. These mechanisms, when paired with our optimisation criterion, structurally bias influencer behaviour toward short-term, polarising strategies. Consequently, our analysis captures the dynamics of extremist content dissemination more than the broader spectrum of influencer behaviours we originally sought to explore.

While the insights into polarisation are valuable, they are ultimately less generalisable to the diverse and adaptive strategies employed by real-world influencers. Future work must assess whether alternative behavioural modes can emerge under different modelling assumptions or whether the current architecture inherently favours extremism.

Additionally, we adopt a scale-based optimisation of influence which excludes the dynamics of micro-influencers. Although less prominent in terms of audience size, micro-influencers have demonstrated outsized influence in political and social discourse (Cha et al., 2010). By conflating influence with scale, our model overlooks these actors and the distinctive mechanisms through which they exert impact.

Another key dimension of real-world influencer success that the model overlooks is relatability and authenticity. Many influencers gain traction not by adopting extreme views, but by cultivating trust, emotional resonance, and a sense of shared identity with their audiences (Borchers, 2025; Duffy, 2020; Prawira et al., 2024). These strategies, often grounded in perceived authenticity rather than novelty, are fundamentally incompatible with the extremity-driven behaviour enforced by our model's optimisation logic.

Another unrealistic assumption in the model is the influencer's ability to accurately assess everyone's true opinion. Human influencers do not possess the computational power or cognitive capacity to analyse the impact of every possible opinion shift. Instead, they rely on personal judgment, experiences, and intuition to gauge what resonates with their audience. This simplification again fails to reflect real influencer strategies, potentially overestimating how successful influencers can be. This limitation was acknowledged in Coculescu et al. (2023), though their proposed solution (a highly abstract mathematical approximation) ultimately reproduces the same unrealistic assumptions.

Moreover, the model assumes that influencers target the entire population, which is an unrealistic portrayal of how influence operates in real-world contexts. In reality, influencers typically focus on a specific target audience and tailor their content to that group, rather than attempting to influence everyone indiscriminately. This issue for our influencer is partly driven by the model's limitation in accounting for the diversity of opinions and personalities in the population. The limited variation in these traits makes it impractical for an influencer to select a target audience for stylistic content despite this being a clear action an influencer would take. Studies such as Coculescu et al. (2023) have explored similar concerns, highlighting the need for a more varied approach to influencer behaviour.

8.4 An Alternative View of The Model

Given the limitations of our current model, we propose an alternative interpretation, one that reimagines the influencer not as a human agent, but as an AI-driven entity. While the model was originally designed to evaluate the role of online content creators, its core assumptions, such as optimisation strategies, fixed networks, and simplified cognitive processes potentially align more naturally with an algorithmic system than with real human behaviour. Since much of our critique focused on the unrealistic portrayal of human influencers, this reinterpretation offers a productive shift: instead of forcing realism onto the influencer, we accept its artificial nature and use it to model a different, but equally relevant, phenomenon.

Such a perspective is particularly valuable in the context of modern social media. AI-driven systems, such as bots or recommendation algorithms on platforms like YouTube or TikTok, continuously monitor user engagement and optimise content delivery in real time. These systems do not rely on intuition or emotion, instead they process massive behavioural datasets to nudge users towards content that maximises some platform objective. In the case of TikTok, for instance, the goal is to keep users scrolling by providing them just enough novelty and emotional stimulus to sustain engagement. This is not unlike our model's influencer, whose sole objective is to maximise influence and attention across the network. Our model similarly assumes an agent with complete access and knowledge to individual opinions, mirroring how real-world AI algorithms or bots operate. No human influencer can achieve this level of granularity or speed but an AI system can, and often does.

An analogy helps clarify this interpretation. Imagine that our influencer node isn't a person at all, but something like TikTok's recommendation engine. It's not implausible to suggest that such a system has considerable control and influence over a group of agents. After all, it determines what content is seen, when, and by whom. Like a recommendation algorithm, it subtly alters the local information available for each agent, not by changing their mind directly, but by changing what surrounds them and what they see as the norm (just like our influencers).

Re framing the model in this way responds to growing calls to understand how AI systems shape political discourse and contribute to ideological polarisation. Previous research shows that algorithms often amplify content not to persuade the majority, but to target and deepen the convictions of minorities (Pariser, 2011). For example, YouTube's recommendation engine has been shown to gradually steer users towards more extreme content over time, reinforcing their initial biases rather than challenging them (Haroon et al., 2022). Similarly, the Facebook–Cambridge Analytica scandal highlighted how detailed profiling and targeting were used to influence voter opinions, again, not by shifting mainstream opinion, but by exploiting vulnerabilities in specific segments of the population (Isaak and Hanna, 2018). The influencer in our model effectively acts as a stand-in for these engines, offering a simplified but conceptually coherent representation of how such systems shape opinion landscapes.

Ultimately, this reinterpretation extends the relevance of our work. While the original framing focused on modelling human influencers, this alternative perspective provides a theoretical entry point into understanding the structural, algorithmic forces that underpin digital polarisation.

We can then investigate the effect real-world algorithmic systems operating under similar incentive structures (e.g. engagement maximisation) will have on populations.

Chapter 9

Future Work

Building on our current model, we propose several avenues for future research to enhance its complexity and applicability:

- **Multiple Influencers:** Pivoting the model to focus on multiple influencers is critical for understanding broader network dynamics. Key questions include: 1) How many influencers are needed to reach a tipping point in the network? 2) How do influencers interact within the same network? Can they mitigate each other's polarising effects and promote balanced discourse, as suggested by Helfmann et al. (2023)?
- **Incorporating Multiple Opinions:** Expanding each agent's opinion set to include multiple dimensions would provide a more nuanced view of opinion evolution across various issues. This approach could capture how lifestyle influencers shift from topics like health and beauty to politics (Borchers, 2025). A multi-opinion framework would enable a deeper analysis of social influence and belief evolution. While discrete opinion models often reduce to single-opinion systems, continuous models introduce greater complexity, as highlighted by Herrerías-Azcué and Galla (2019), thus calling for further research.
- **Permanent Agent Characteristics:** Introducing permanent agent attributes (e.g., age, gender, socioeconomic status) could refine our understanding of how homophily influences social interactions. This would better reflect the role of demographic factors in polarisation, such as the rise of far-right views among young men (Lee, 2024; Nilan et al., 2023), and explain how shared identities affect influencer dynamics.
- **Exploring Influencer Strategy Variation:** Expanding influencer strategies beyond the extremism driven by computational constraints in the current model would provide deeper insights into their role in opinion dynamics. Influencers employ diverse strategies, as documented in studies such as (Goodwin et al., 2023a; Hegselmann et al., 2015). Incorporating these variations could involve hard coding different strategies based on common tactics observed in online platforms, overriding the computational limitations imposed by the model's edge update.
- **Heterogeneous Parameters:** Introducing variability in parameters such as θ_a and θ_h across agents would better capture the diversity of social behaviours in real networks. This would allow for a more realistic modelling of how attention to novelty and homophily vary among individuals.

Chapter 10

Conclusion

In this work, we have developed a comprehensive framework for analysing the impact of influencers on opinion dynamics. We reviewed and critiqued existing models, highlighting the cultural tendencies of the field, while positioning our model as a step towards a more empirically focused perspective. Regarding the aims outlined in Chapter 1, we believe all objectives were achieved. We discuss a minor criticism of objective (3) below, acknowledging that while it enabled strategic freedom, it also introduced certain problems.

Our model guided influencer strategy through the optimisation of out-strength; however, these strategies were fundamentally constrained by the structural and dynamical properties of the system. Notably, the interaction mechanisms embedded within the base model facilitated strategic extremism as a consistently effective means for influencers to increase their prominence. This raises a critical question for future work: whether alternative, non-extremist strategies can emerge as successful under the same optimisation criterion, or whether the architecture of the model inherently biases the system towards extremism. Addressing this will be essential for evaluating a greater variety of influencer mechanisms within this model.

Our results underscore the complex dynamics of polarisation. While a single influencer introduced minimal disruption, the presence of multiple influencers triggered widespread polarisation. Although real polarisation can enhance democratic engagement, by sharpening ideological distinctions and encouraging participation, its benefits rely on sustained open dialogue (Fiorina, 2010; Vasconcelos et al., 2021). In practice, it often hardens divisions, reconfigures personal identities and social networks, and shifts political discourse towards antagonism over substance (Axelrod et al., 2021; Bail, Argyle, Brown, and Volfovsky, 2018; Iyengar and Westwood, 2014; McCoy, 2019). The emergence of this dynamic highlights the critical role influential agents play in deepening societal divides. We hope this encourages future studies to incorporate more empirical investigations of influential agents, particularly through social media analysis, to identify such behaviours.

In examining vulnerable agents, we identified a key mechanism through which extremism propagates in opinion networks. Specifically, we found that agents with low homophily and high sensitivity to novelty are disproportionately susceptible to influence. Influencers, arguably without explicit intent, end up exploiting these traits, shifting local norms by disproportionately affecting the most susceptible nodes. The structural dynamics of the model result in these agents becoming the primary conduits for influence. To our knowledge, no prior work on influencers has explicitly addressed this dynamic. This challenges the common assumption of uniform spread in opinion dynamics. Instead, highlighting specific cognitive configurations as leverage points for the diffusion of polarising content. We showed how this aligned with key empirical findings in the real-world.

Finally, we aimed to ground our technical decisions in empirical justification wherever possible. However, we acknowledge that greater empirical validation is necessary to robustly corroborate our findings. While our model captures key dynamics observed in real-world opinion formation, such as the influence of extremism and the vulnerability of certain agent profiles, its generalisability relies on further empirical testing.

Ultimately, this study contributes to a deeper understanding of how strategic behaviour by influencers, particularly through differentiation, can reshape the ideological evolution of social systems. As the digital public sphere grows more fragmented and algorithmically curated, our models must evolve in parallel, capable of capturing not just the complexity of targeted influence, but its capacity to entrench division, distort norms, and steer collective belief.

Appendix A

Appendix A: Additional Measures

We include various tables summarising statistical tests performed during the project.

Test	Brown-Forsythe	Variance Equality	T-test	Significance
Mean Edge Weight	0.63	equal variances	0.511	not different means
Assortativity	0.29	equal variances	0.269	not different means
Modularity	0.3	equal variances	0.33	not different means
Community Std. Dev. of Opinion	0.39	equal variances	0.58	not different means
Community Std. Dev. of Size	0.02	unequal variances	1.05E-06	different means
Node Deviance	0.51	equal variances	0.387	not different means

Table A.1: T-test results comparing network measures between influencer and non-influencer networks of n = 1000 across 30 simulations.

Test	Brown-Forsythe	Variance Equality	T-test	Significance
Conformity (c)	5.40E-08	different variances	1.7E-18	different means
Attention to Novelty (a)	0.0	different variances	0.0	different means
Homophily (h)	0.0	different variances	0.0	different means

Table A.2: T-test results comparing c, h, a values between average and vulnerable agents.

Test	Correlation	P-Value
6.5a	0.1960	0.2991
6.5b	0.1521	0.4223
6.5c	0.0523	0.7833

Table A.3: T-test results comparing c, h, a values between average and vulnerable agents.

Test	Brown-Forsythe	Variance Equality	T-test	Significance
Mean Edge Weight	0.0	different variances	0.0	different means
Assortativity	0.09	equal variances	3.03E-14	different means
Modularity	0.0165	different variances	2.01E-16	different means
Community Std. Dev. of Opinino	0.015	different variances	2.1E-11	different means
Community Std. Dev. of Size	0.077	equal variances	1.9E-16	different means
Node Deviance	0.011	different variances	1.61E-11	different means

Table A.4: T-test results comparing network measures between homogeneous and heterogeneous networks of n = 1000 across 5 simulations.

Parameter	Cohen's d
Homophily	-1.346
Attention to Novelty	1.346
Conformity	0.229

Table A.5: Cohen's d values for top 50 neighbours vs network average values. The effect sizes quantify just how different the c, h and a values are for our neighbours in relation to the network.

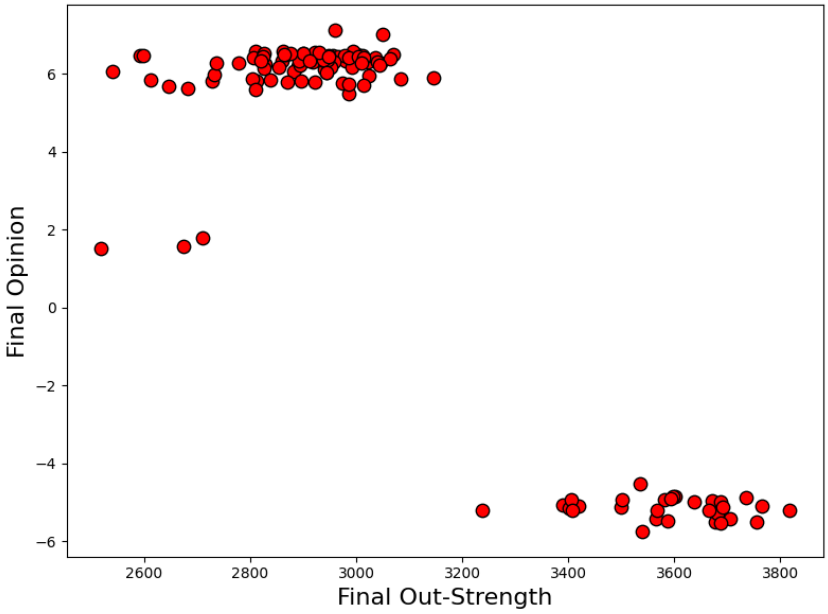


Figure A.1: Plot of final opinion vs. final out-strength for 100 influencers in a network. The out-strength distribution is skewed, reflecting the initial positive bias in opinions. Notably, the final opinion stabilises within the $[6, -6]$ range, rather than the expected $[4, -4]$, due to the reduced neutral population and increased polarisation.

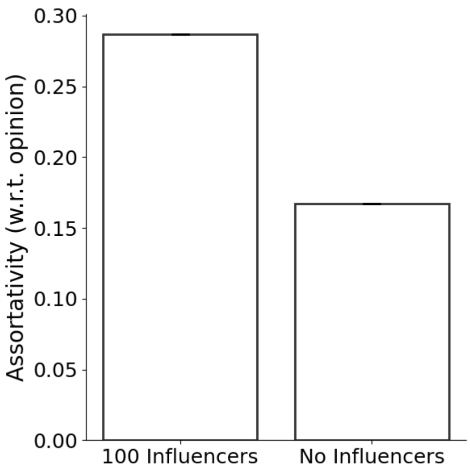


Figure A.2: Comparison of assortativity for networks with and without 100 influencers. There is a considerably higher assortativity for multiple influencer networks.

Appendix B

Appendix B: Artificial Intelligence

BibTex Automation

Many research websites and academic journals provide options to export citations in BibTeX format. While some offer a straightforward copy-and-paste feature, others require downloading a citation file and importing it into a LaTeX editor. To streamline this process, I used ChatGPT to generate BibTeX entries directly. By pasting the citation text into the prompt with a query such as *bibtex cite this*, I could quickly obtain a correctly formatted BibTeX entry. I then cross-verified the output against the original source to ensure accuracy, eliminating the need for repetitive downloading, file management, and cleanup.

Spelling and Grammar

On my final read-throughs, I used ChatGPT to check for spelling and/or grammar checks. This was applied solely to Chapters 2 and 10 with the prompt being for the background being: *any issues with punctuation and spag, focus on the dashes, and british grammar/spag*. For the conclusion: *any spelling and grammar and punctuation issues?*. This was because I'd noticed a mis-match in the types of dashes I was using within these sections.

LaTex Formatting

For data tables, I collated data in Microsoft Excel, typically from command line output, and then used ChatGPT to format this into a LaTeX-friendly table. This was a simple prompt line of: *format into a latex table*.

Bibliography

Allport, G. W. (1954). *The nature of prejudice*. Addison-Wesley.

Antoniades, D., & Dovrolis, C. (2015). Co-evolutionary dynamics in social networks: A case study of twitter. *Computational Social Networks*, *2*(1), 14.

Asch, S. (1956). Studies of independence and conformity: I. a minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, *70*. https://doi.org/10.1037/h0093718

Axelrod, R., Daymude, J. J., & Forrest, S. (2021). Preventing extreme polarization of political attitudes. *Proceedings of the National Academy of Sciences*, *118*(50), e2102139118. https://doi.org/10.1073/pnas.2102139118

Bacon, F. (1620). *Novum organum* (T. Fowler, Ed.) [Original work published in 1620]. Clarendon Press.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, *115*(37), 9216–9221. https://doi.org/10.1073/pnas.1804840115

Bail, C. A., Argyle, L. P., Brown, T. W., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, *115*(37), 9216–9221. https://doi.org/10.1073/pnas.1804840115

Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Everyone's an influencer: Quantifying influence on twitter. *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, 65–74. https://doi.org/10.1145/1935826.1935845

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512. https://doi.org/10.1126/science.286.5439.509

Baumann, F., Lorenz-Spreen, P., Sokolov, I. M., & Starnini, M. (2020). Modeling echo chambers and polarization dynamics in social networks. *Phys. Rev. Lett.*, *124*, 048301. https://doi.org/10.1103/PhysRevLett.124.048301

Baumann, F., Sokolov, I. M., & Tyloo, M. (2020). A laplacian approach to stubborn agents and their role in opinion formation on influence networks. *Physica A: Statistical Mechanics and its Applications*, *557*, 124869. https://doi.org/10.1016/j.physa.2020.124869

Behera, L., & Schweitzer, F. (2003). On spatial consensus formation: Is the snazjd model different from a voter model? *International Journal of Modern Physics C*, *14*(10), 1331–1354. https://doi.org/10.1142/s0129183103005467

Bernardes, A. T., Costa, U. M. S., Araujo, A. D., & Stauffer, D. (2001). Damage Spreading, Coarsening Dynamics And Distribution Of Political Votes In Sznajd Model On Square Lattice. *International Journal of Modern Physics C (IJMPC)*, *12*(02), 159–167. https://doi.org/10.1142/S0129183101001584

Berner, R., Yanchuk, S., & Schöll, E. (2021). What adaptive neuronal networks teach us about power grids. *Phys. Rev. E*, *103*, 042315. https://doi.org/10.1103/PhysRevE.103.042315

Betts, J. M., & Bliuc, A.-M. The effect of influencers on societal polarization [Cited by: 1; All Open Access, Green Open Access]. In: *2022-December*. Cited by: 1; All Open Access, Green Open Access. 2022, 370–381. https://doi.org/10.1109/WSC57314.2022.10015491

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Bogardus, E. S. (1924). *Fundamentals of social psychology*. Century.

Borchers, N. S. (2025). How social media influencers support political parties in achieving campaign objectives, according to political communicators in germany. *Public Relations Review*, *51*(1), 102532. https://doi.org/https://doi.org/10.1016/j.pubrev.2024.102532

Bouchaud, J.-P., Borghesi, C., & Jensen, P. (2014). On the emergence of an 'intention field' for socially cohesive agents. *J. Stat. Mech.: Theory Exp.*, *2014*, P03010. https://doi.org/10.1088/1742-5468/2014/03/P03010

Brooks, H. Z., & Porter, M. A. (2020). A model for the influence of media on the ideology of content in online social networks. *Phys. Rev. Res.*, *2*, 023041. https://doi.org/10.1103/PhysRevResearch.2.023041

Brown, S. (1996). Public opinion and the communication of consent theodore l. glasser and charles t. salmon, eds. new york: Guilford press, 1995, pp. xxxiv, 475. *Canadian Journal of Political Science*, *29*, 592–593. https://doi.org/10.1017/S0008423900008386

Bullock, S., & Sayama, H. (2023, May). Agent heterogeneity mediates extremism in an adaptive social network model [The 2023 Conference on Artificial Life ; Conference date: 24-07-2023 Through 28-07-2023]. In H. Iizuka, K. Suzuki, R. Uno, L. Damiano, N. Spychalav, M. Aguilera, E. Izquierdo, R. Suzuki, & M. Baltieri (Eds.), *Proceedings of the artificial life conference 2023 (alife 2023)* (pp. 258–266). Massachusetts Institute of Technology (MIT) Press. https://doi.org/10.1162/isal_a_00628

Cao, W., Zhang, H., Kou, G., & Zhang, B. (2024). Discrete opinion dynamics in social networks with stubborn agents and limited information. *Information Fusion*, *109*, 102410. https://doi.org/https://doi.org/10.1016/j.inffus.2024.102410

Carpentras, D. (2023). Why we are failing at connecting opinion dynamics to the empirical world [ETH Zürich – Department of Humanities, Social and Political Sciences (GESS)]. *The Submission Author*. https://rofasss.org/2023/03/08/od-emprics/#comment-114952

Carpentras, D., Maher, P. J., O'Reilly, C., & Quayle, M. (2022). Deriving an opinion dynamics model from experimental data. *Journal of Artificial Societies and Social Simulation*, *25*(4), 4. https://doi.org/10.18564/jasss.4947

Castellano, C., Vilone, D., & Vespignani, A. (2003). Incomplete ordering of the voter model on small-world networks. *Europhysics Letters*, *63*(1), 153. https://doi.org/10.1209/epl/i2003-00490-0

Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. *Rev. Mod. Phys.*, *81*, 591–646. https://doi.org/10.1103/RevModPhys.81.591

Castellano, C., Muñoz, M., & Pastor-Satorras, R. (2009). Nonlinear q -voter model. *Physical review. E, Statistical, nonlinear, and soft matter physics*, *80*, 041129. https://doi.org/10.1103/PhysRevE.80.041129

Centola, D., Becker, J., Brackbill, D., & Baronchelli, A. (2018). Experimental evidence for tipping points in social convention. *Science*, *360*(6393), 1116–1119. https://doi.org/10.1126/science.aas8827

Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. P. (2010). Measuring user influence in twitter: The million follower fallacy. *AAAI Conference on Weblogs and Social Media*, *14*.

Chacoma, A., & Zanette, D. H. (2015). Opinion formation by social influence: From experiments to modeling. *PLoS ONE*, *10*(10), e0140406. https://doi.org/10.1371/journal.pone.0140406

Chattoe-Brown, E. (2022, February). If you want to be cited, don't validate your agent-based model: A tentative hypothesis badly in need of refutation [Accessed April 18, 2025]. https://rofasss.org/2022/02/01/citing-od-models/

Chazelle, B. (2012). The dynamics of influence systems. *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, 311–320. https://doi.org/10.1109/FOCS.2012.70

Chen, K., Luo, Y., Hu, A., & Liwei, Z. (2021). Characteristics of misinformation spreading on social media during the covid-19 outbreak in china: A descriptive analysis. *Risk Management and Healthcare Policy*, *14*, 1869–1879. https://doi.org/10.2147/RMHP.S312327

Cheng, Z., Chen, J., Peng, R., & Shoenberger, H. (2023). Social media influencers talk about politics: Investigating the role of source factors and psr in gen-z followers' perceived information quality, receptivity and sharing intention. *Journal of Information Technology Politics*, *21*. https://doi.org/10.1080/19331681.2023.2173700

Clifford, P., & Sudbury, A. (1973). A model for spatial conflict. *Biometrika*, *60*(3), 581–588. Retrieved March 2, 2025, from http://www.jstor.org/stable/2335008

Coculescu, D., Motte, M., & Pham, H. (2023). Opinion dynamics in communities with major influencers and implicit social influence via mean-field approximation. https://arxiv.org/abs/2306.16553

Deffuant, G., Neau, D., Amblard, F., & Weisbuch, G. (2000). Mixing beliefs among interacting agents. *Advances in Complex Systems*, *03*(01n04), 87–98. https://doi.org/10.1142/S0219525900000078

Degroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, *69*, 118–121. https://api.semanticscholar.org/CorpusID:120077293 den Heijer, M., Jansen, S., Pellemans, M., & Treur, J. (2024). A second-order adaptive network model for political opinion dynamics.

Ding, C., Jabr, W., & Guo, H. (2023). Electoral competition in the age of social media: The role of social media influencers [Published Online: November 30, 2023, Published in Issue: December 1, 2023]. *MIS Quarterly*, *47*(4), 1647–1662. https://doi.org/10.25300/MISQ/2022/16422

Doherty, M. E., & Mynatt, C. R. (1987). The magical number one. In D. R. Moates & R. Butrick (Eds.), *Proceedings of the ohio university interdisciplinary inference conference* (pp. 221–230).

Dong, J., Zhang, Y.-C., & Kong, Y. (2024). The evolution dynamics of collective and individual opinions in social networks. *Expert Systems with Applications*, *255*, 124813. https://doi.org/https://doi.org/10.1016/j.eswa.2024.124813

Dong, Y., Zha, Q., Zhang, H., & Herrera, F. (2020). Consensus reaching and strategic manipulation in group decision making with trust relationships. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *PP*, 1–15. https://doi.org/10.1109/TSMC.2019.2961752

Dorogovtsev, S. N., Goltsev, A. V., & Mendes, J. F. F. (2008). Critical phenomena in complex networks. *Reviews of Modern Physics*, *80*(4), 1275–1335. https://doi.org/10.1103/revmodphys.80.1275

Duffy, B. E. (2020). Social media influencers. In K. Ross, F. Attwood, V. Cardo, J. Mercer, & S. Moorti (Eds.), *Encyclopedia of gender, media, and communication* (pp. 1–5). Wiley-Blackwell. https://doi.org/10.1002/9781119429128.iegmc219

Enke, N., & Borchers, N. (2019). Social media influencers in strategic communication: A conceptual framework for strategic social media influencer communication. *International Journal of Strategic Communication*, *13*, 261–277. https://doi.org/10.1080/1553118X.2019.1620234

Euler, L. (1736). Solutio problematis ad geometriam situs pertinentis [English translation: *The solution of a problem relating to the geometry of position*]. *Commentarii academiae scientiarum Petropolitanae*, *8*, 128–140.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

Fiorina, M. P. (2010). Why political polarization might be good for america. *US News & World Report*. https://www.usnews.com/opinion/articles/2010/05/27/why-political-polarization-might-be-good-for-america

Fischer, T.-S., Kolo, C., & Mothes, C. (2022). Political influencers on youtube: Business strategies and content characteristics. *Media and Communication*, *10*(1), 259–271. https://doi.org/10.17645/mac.v10i1.4767

Flamino, J., Galeazzi, A., Feldman, S., Macy, M., Cross, B., Zhou, Z., Serafino, M., Bovet, A., Makse, H., & Szymanski, B. (2023). Political polarization of news media and influencers on twitter in the 2016 and 2020 us presidential elections. *Nature Human Behaviour*, *7*(6), 904–916. https://doi.org/10.1038/s41562-023-01550-8

Fortunato, S. (2009). Community detection in graphs. *Physics Reports*, *486*. https://doi.org/10.1016/j.physrep.2009.11.002

French, J., & Raven, B. (1959, January). *The bases of social power* (Vol. 6).

Frey, D. (1986). Recent research on selective exposure to information. *Advances in Experimental Social Psychology*, *19*, 41–80. https://doi.org/10.1016/S0065-2601(08)60212-9

Gentzkow, M., Shapiro, J. M., & Stone, R. (2018). Media bias and polarization in the age of social media. *Journal of Economic Perspectives*, *32*(3), 107–130.

Gentzkow, M., & Shapiro, J. M. (2018). How tech platforms fuel u.s. political polarization and what government can do about it [Accessed: 2025-04-22]. https://www.brookings.edu/articles/how-tech-platforms-fuel-u-s-political-polarization-and-what-government-can-do-about-it/

Glazer, E., & Wells, G. (2019). Political campaigns turn to social media influencers to reach voters [Accessed: 2025-04-21]. *The Wall Street Journal*. https://www.wsj.com/articles/political-campaigns-turn-to-social-media-influencers-to-reach-voters-11569251450

Gleeson, J. P. (2011). High-accuracy approximation of binary-state dynamics on networks. *Phys. Rev. Lett.*, *107*, 068701. https://doi.org/10.1103/PhysRevLett.107.068701

Goldenberg, A., Bailey, D., Murić, G., Ferrara, E., Schöne, J., Willer, R., Halperin, E., & Gross, J. (2023, March). *Attraction to politically extreme users on social media*. https://doi.org/10.31219/osf.io/cmx4p

Goodwin, A., Joseff, K., Riedl, M., Lukito, J., & Woolley, S. (2023a). Political relational influencers: The mobilization of social media influencers in the political arena. *International Journal of Communication*, *17*(0). https://ijoc.org/index.php/ijoc/article/view/18987

Goodwin, A., Joseff, K., Riedl, M., Lukito, J., & Woolley, S. (2023b). Political relational influencers: The mobilization of social media influencers in the political arena. *International Journal of Communication*, *17*(0). https://ijoc.org/index.php/ijoc/article/view/18987

Granovsky, B. L., & Madras, N. (1995). The noisy voter model. *Stochastic Processes and their Applications*, *55*(1), 23–43. https://ideas.repec.org/a/eee/spapps/v55y1995i1p23-43.html

Gross, T., & Blasius, B. (2007). Adaptive coevolutionary networks: A review. *Journal of The Royal Society Interface*, *5*(20), 259–271. https://doi.org/10.1098/rsif.2007.1229

Ha, L. (2024). Behind the screens: Insights from digital content creators; understanding their intentions, practices and challenges. https://unesdoc.unesco.org/ark:/48223/pf0000392006

Haney, C., Banks, C., & Zimbardo, P. (1973). Interpersonal dynamics in a simulated prison. *International Journal of Criminology and Penology*, *1*, 69–97.

Harff, D., & Schmuck, D. (2023). Influencers as empowering agents? following political influencers, internal political efficacy and participation among youth. *Political Communication*, *40*, 1–26. https://doi.org/10.1080/10584609.2023.2166631

Haroon, M., Chhabra, A., Liu, X., Mohapatra, P., Shafiq, Z., & Wojcieszak, M. (2022). Youtube, the great radicalizer? auditing and mitigating ideological biases in youtube recommendations. https://arxiv.org/abs/2203.10666

Hassan, G., Brouillette-Alarie, S., Alava, S., Sieckelinck, S., et al. (2018). Exposure to extremist online content could lead to violent radicalization: A systematic review of empirical evidence. *International Journal of Developmental Science*, *12*(7), 1–18. https://doi.org/10.3233/DEV-170233

Hegselmann, R., König, S., Kurz, S., Niemann, C., & Rambau, J. (2015). Optimal opinion control: The campaign problem. https://arxiv.org/abs/1410.8419

Hegselmann, R., & Krause, U. (2002). Opinion dynamics and bounded confidence: Models, analysis and simulation. *Journal of Artificial Societies and Social Simulation*, *5*.

Hegselmann, R., & Krause, U. (2015). *Opinion dynamics under the influence of radical groups, charismatic leaders, and other constant signals: A simple unifying model*. https://doi.org/10.3934/nhm.2015.10.477

Helfmann, L., Conrad, N. D., Lorenz-Spreen, P., & Schütte, C. (2023). Modelling opinion dynamics under the impact of influencer and media strategies. https://arxiv.org/abs/2301.13661

Herrerías-Azcué, F., & Galla, T. (2019). Consensus and diversity in multistate noisy voter models. *Phys. Rev. E*, *100*, 022304. https://doi.org/10.1103/PhysRevE.100.022304

Hofmann, S. G., & Curtiss, J. (2018). A complex network approach to clinical science. *Eur. J. Clin. Invest.*, *48*(8), e12986.

Holley, R. A., & Liggett, T. M. (1975). Ergodic Theorems for Weakly Interacting Infinite Systems and the Voter Model. *The Annals of Probability*, *3*(4), 643–663. https://doi.org/10.1214/aop/1176996306

Hołyst, J., & Kacperski, K. (2001). Social impact models of opinion dynamics. *Ann Rev Comp Phys*, *9*. https://doi.org/10.1142/9789812811578_0005

Homans, G. C. (1958). Social behavior as exchange. *American Journal of Sociology*, *63*(6), 597–606. Retrieved April 1, 2025, from http://www.jstor.org/stable/2772990

Horgan, J. (2017). *The psychology of terrorism*. Routledge.

Hu, H., Chen, W., & Hu, Y. (2024). Opinion dynamics in social networks under the influence of mass media. *Applied Mathematics and Computation*, *482*, 128976. https://doi.org/https://doi.org/10.1016/j.amc.2024.128976

Hugh, D., Dolan, R., Harrigan, P., & Gray, H. (2022). Influencer marketing effectiveness: The mechanisms that matter. *European Journal of Marketing*, *56*. https://doi.org/10.1108/EJM-09-2020-0703

Huszár, F., Ktena, S. I., O'Brien, C., Belli, L., Schlaikjer, A., & Hardt, M. (2022). Algorithmic amplification of politics on twitter. *Proceedings of the National Academy of Sciences*, *119*(1), e2025334119. https://doi.org/10.1073/pnas.2025334119

Huynh, H. N., Ng, K. L., Toh, R., & Feng, L. (2024). Understanding the impact of network structure on air travel pattern at different scales. *PLOS ONE*, *19*(3), 1–20. https://doi.org/10.1371/journal.pone.0299897

Isaak, J., & Hanna, M. (2018). User data privacy: Facebook, cambridge analytica, and privacy protection. *Computer*, *51*, 56–59. https://doi.org/10.1109/MC.2018.3191268

Iyengar, S., & Westwood, S. J. (2014). Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science*, *59*(3), 690–707. https://doi.org/10.1111/ajps.12152

Judijanto, L., Harlina, S., Ar, M., Desak, N., Diwyarthi, M., Setiadi, B., Bali, P., Tinggi, S., & Aviasi, P. (2024). The influence of social media influencer utilization in influencing public opinion. *West Science Social and Humanities Studies*, *Vol. 02*, 40–49. https://doi.org/10.58812/wsshs.v2i01.548

Kaizoji, T. (2006). An interacting-agent model of financial markets from the viewpoint of nonextensive statistical mechanics [Econophysics Colloquium]. *Physica A: Statistical Mechanics and its Applications*, *370*(1), 109–113. https://doi.org/https://doi.org/10.1016/j.physa.2006.04.031

Kaizoji, T., Bornholdt, S., & Fujiwara, Y. (2002). Dynamics of price and trading volume in a spin model of stock markets with heterogeneous agents. *Physica A: Statistical Mechanics and its Applications*, *316*(1), 441–452. https://doi.org/https://doi.org/10.1016/S0378-4371(02)01216-5

Katz, E., & Lazarsfeld, P. (1955). *Personal influence: The part played by people in the flow of mass communications*. Free Press.

Kelley, H. H., & Volkart, E. H. (1952). The resistance to change of group-anchored attitudes. *American Sociological Review*, *17*(4), 453–465. Retrieved April 15, 2025, from http://www.jstor.org/stable/2088001

Keynes, J. M. (1936). *The general theory of employment, interest and money*. Macmillan.

Klapper, J. T. (1960). *The effects of mass communication*. Free Press.

Kosse, C., & Carstens, S. (2023). The role of influencers in driving political polarization: The effects of extremity and follower engagement. *Journal of Digital Politics*, *5*(2), 45–63.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J. H., & Gutmann, M. (2009). Life in the network: The coming age of computational social science. *Science*, *323*(5915), 721–723. https://doi.org/10.1126/science.1167742

Lee, S. C. (2024). Anti-gender politics, economic insecurity, and right-wing populism: The rise of modern sexism among young men in south korea. *Social Politics: International Studies in Gender, State Society*, jxae016. https://doi.org/10.1093/sp/jxae016

Levy, R. (2021). Social media, news consumption, and polarization: Evidence from a field experiment. *American Economic Review*, *111*, 831–870. https://doi.org/10.1257/aer.20191777

Lian, Y., & Dong, X. (2022). An opinion dynamics model for unrelated discrete opinions. *Knowledge-Based Systems*, *251*, 109133. https://doi.org/https://doi.org/10.1016/j.knosys.2022.109133

Liu, B. (2012). *Sentiment analysis and opinion mining* (Vol. 5). Morgan & Claypool Publishers. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Marconi, L., & Cecconi, F. (2020). Opinion dynamics and consensus formation in a deffuant model with extremists and moderates. https://arxiv.org/abs/2010.01534

Marriott, J. (2024). How joe rogan helped donald trump to win america [Accessed: [Insert date you accessed the article]]. *The Times*.

Martins, A. C., Pereira, C. d. B., & Vicente, R. (2009). An opinion dynamics model for the diffusion of innovations. *Physica A: Statistical Mechanics and its Applications*, *388*(15–16), 3225–3232. https://doi.org/10.1016/j.physa.2009.04.007

Mason, W. A., Conrey, F. R., & Smith, E. R. (2007). Situating social influence processes: Dynamic, multidirectional flows of influence within social networks [PMID: 18453465]. *Personality and Social Psychology Review*, *11*(3), 279–300. https://doi.org/10.1177/1088868307301032

McCauley, C., & Moskalenko, S. (2008). Mechanisms of political radicalization: Pathways toward terrorism. *Terrorism and Political Violence*, *20*(3), 415–433. https://doi.org/10.1080/09546550802073317

McCoy, J. (2019). Polarization harms democracy and society. *Peace in Progress*, (36). https://www.icip.cat/perlapau/en/article/polarization-harms-democracy-and-society/

McPherson, M., Smith-Lovin, L., & Cook, J. (2001). Birds of a feather: Homophily in social networks. *Annu Rev Sociol*, *27*, 415–444. https://doi.org/10.1146/annurev.soc.27.1.415

Milgram, S. (1964). Issues in the study of obedience: A reply to baumrind. *American Psychologist*, *19*, 848–852. https://doi.org/10.1037/h0044954

Mobilia, M., Petersen, A., & Redner, S. (2007). On the role of zealotry in the voter model. *Journal of Statistical Mechanics: Theory and Experiment*, *2007*(08), P08029–P08029. https://doi.org/10.1088/1742-5468/2007/08/p08029

Moscovici, S., & Nève, P. (1971). Studies in social influence: I. those absent are in the right: Convergence and polarization of answers in the course of a social interaction. *European Journal of Social Psychology*, *1*(2), 201–214. https://doi.org/https://doi.org/10.1002/ejsp.2420010204

Motsch, S., & Tadmor, E. (2014). Heterophilious dynamics enhances consensus. https://arxiv.org/abs/1301.4123

Moussaïd, M., Kämmer, J. E., Analytis, P. P., & Neth, H. (2013). Social influence and the collective dynamics of opinion formation. *PLoS ONE*, *8*(11), e78433. https://doi.org/10.1371/journal.pone.0078433

Muñoz, P., Bellogín, A., Barba-Rojas, R., & Díez, F. (2024). Quantifying polarization in online political discourse [Accessed: 2025-04-22]. *EPJ Data Science*, *13*(1), 39. https://doi.org/10.1140/epjds/s13688-024-00480-3

Nguyen, V., & Shvydkoy, R. (2023). *Continuous model of opinion dynamics with convictions*. https://doi.org/10.3934/dcds.2023076

Nickerson, R. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Nilan, P., Roose, J., Peucker, M., & Turner, B. S. (2023). Young masculinities and right-wing populism in australia. *Youth*, *3*(1), 285–299. https://doi.org/10.3390/youth3010019

Nöfer, E. (2024). Disinformation and the role of social media influencers in times of crises, conflicts, and wars: Recommendations on countering disinformation [Accessed: 2025-04-23]. https://www.aspeninstitute.de/wp-content/uploads/Disinformation-and-the-Role-of-Social-Media-Influencers-in-Times-of-Crises-Conflicts-and-Wars-Recommendations-on-Countering-Disinformation-2.pdf

Osnos, E. (2019). When gun violence meets ideology. *The New Yorker*. https://www.newyorker.com/news/news-desk/when-gun-violence-meets-ideology

Panizza, F., Vostroknutov, A., & Coricelli, G. (2021). How conformity can lead to polarised social behaviour. *PLoS Comput. Biol.*, *17*(10), e1009530.

Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin Group , The.

Peralta, A. F., Kertész, J., & Iñiguez, G. (2022). Opinion dynamics in social networks: From models to data. https://arxiv.org/abs/2201.01322

Perra, N., & Rocha, L. E. C. (2018). Modelling opinion dynamics in the age of algorithmic personalisation. https://arxiv.org/abs/1811.03341

Perrin, A. (2015). *Social media usage: 2005-2015: 65% of adults now use social networking sites–a nearly tenfold jump in the past decade*. Pew Research Trust. https://books.google.co.uk/books?id=OupAnQAACAAJ

Prawira, Y., Sumartias, S., Suryana, A., & Damayani, N. (2024). Political influencers in online political participation: Strategies and challenges on social media platforms. *International Journal of Religion*, *5*, 5485–5494. https://doi.org/10.61707/wz4a1g03

Prior, M. (2013). Media and political polarization [First published as a Review in Advance on February 01, 2013]. *Annual Review of Political Science*, *16*, 101–127. https://doi.org/10.1146/annurev-polisci-100711-135242

Proulx, S. R., Promislow, D. E., & Phillips, P. C. (2005). Network thinking in ecology and evolution [SPECIAL ISSUE: BUMPER BOOK REVIEW]. *Trends in Ecology Evolution*, *20*(6), 345–353. https://doi.org/https://doi.org/10.1016/j.tree.2005.04.004

Ramachandran, V. S. (2010). Carried to extremes. *Scientific American*, *303*(1), 62–69. https://www.scientificamerican.com/article/carried-to-extremes/

Rashotte, L. (2007). Social influences. In G. Ritzer (Ed.), *The blackwell encyclopedia of sociology*. Blackwell Publishing. https://doi.org/10.1002/9781405165518.wbeoss154

Rodrigo, A. G. (2025). Extending the hegselmann-krause model of opinion dynamics to include ai oracles. https://arxiv.org/abs/2502.19701

Safari-Alighiarloo, N., Taghizadeh, M., Rezaei-Tavirani, M., Goliaei, B., & Peyvandi, A. A. (2014). Protein-protein interaction networks (PPI) and complex diseases. *Gastroenterol. Hepatol. Bed Bench*, *7*(1), 17–31.

Sayama, H. (2020). Extreme ideas emerging from social conformity and homophily: An adaptive social network model, 113–120. https://doi.org/10.1162/isal_a_00349

Sayama, H., Pestov, I., Schmidt, J., Bush, B. J., Wong, C., Yamanoi, J., & Gross, T. (2013). Modeling complex systems with adaptive networks [Grasping Complexity]. *Computers Mathematics with Applications*, *65*(10), 1645–1664. https://doi.org/https://doi.org/10.1016/j.camwa.2012.12.005

Schleffer, G., & Miller, B. (2021). The political effects of social media platforms on different regime types. *Texas National Security Review*, *4*(3), 1–18. https://tnsr.org

Seeger, C., & Muth, L. (2023). Social media influencers' role in shaping political opinions and actions of young audiences. *Media and Communication*, *11*. https://doi.org/10.17645/mac.v11i3.6750

Sherif, M. (1936). *The psychology of social norms*. Harper Brothers.

Simi, P., & Futrell, R. (2010). *American swastika: Inside the white power movement's hidden spaces*. University of California Press.

Sîrbu, A., Loreto, V., Servedio, V. D. P., & Tria, F. (2016, May). Opinion dynamics: Models, extensions and external effects. In *Participatory sensing, opinions and collective awareness* (pp. 363–401). Springer International Publishing. https://doi.org/10.1007/978-3-319-25658-0_17

Soares, F. B., Recuero, R., & Zago, G. (2018). Influencers in polarized political networks on twitter. *Proceedings of the 9th International Conference on Social Media and Society*, 168–177. https://doi.org/10.1145/3217804.3217909

Sobkowicz, P. (2009)). Modelling opinion formation with physics tools: Call for closer link with reality. *Journal of Artificial Societies and Social Simulation*, *12*(1), 11. https://www.jasss.org/12/1/11.html

Sood, V., & Redner, S. (2005). Voter model on heterogeneous graphs. *Phys. Rev. Lett.*, *94*, 178701. https://doi.org/10.1103/PhysRevLett.94.178701

Stafford, M. C., & Ekland-Olson, S. (1982). On social learning and deviant behavior: A reappraisal of the findings. *American Sociological Review*, *47*(1), 167–169. Retrieved April 19, 2025, from http://www.jstor.org/stable/2095053

Stasser, G., & Titus, W. (1985). Pooling unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology*, *48*(6), 1467–1478. https://doi.org/10.1037/0022-3514.48.6.1467

Stauffer, D. (2002). The sznajd model of consensus building with limited persuasion. *International Journal of Modern Physics C*, *13*(03), 315–317. https://doi.org/10.1142/s0129183102003140

Stewart, B., Wenzlaff, R., & Cordero, C. (2024). *Influencers get access to political conventions as campaigns court young voters* [Accessed: 2024-11-13]. ABC News. https://abcnews.go.com/Politics/influencers-access-political-conventionscampaigns-court-young-voters/story?id=113100221

Sunstein, C. (2007). Ideological amplification. *Constellations*, *14*, 273–279. https://doi.org/10.1111/j.1467-8675.2007.00439.x

Sureda, C., Gaudou, B., & Amblard, F. (2017). An agent-based simulation of extremist network formation through radical behavior diffusion, 236–243. https://doi.org/10.5220/0006198602360243

Sznajd-Weron, K., & Sznajd, J. (2000). Opinion evolution in closed community. *International Journal of Modern Physics C*, *11*(06), 1157–1165. https://doi.org/10.1142/s0129183100000936

Turner, J. (1985). Social categorization and the self-concept: A cosial cognitive theory of group behavior. *Advances in Group Processes*, *2*.

Turner, J. (1991). *Social influence*. Open University Press.

Vande Kerckhove, C., Martin, S., Gend, P., Rentfrow, P. J., Hendrickx, J. M., & Blondel, V. D. (2016). Modelling influence and opinion evolution in online collective behaviour. *PLoS ONE*, *11*(6), e0157685. https://doi.org/10.1371/journal.pone.0157685

Vasconcelos, V. V., Constantino, S. M., Dannenberg, A., & Levin, S. (2021). Segregation and clustering of preferences erode socially beneficial coordination. *Proceedings of the National Academy of Sciences*, *118*(50), e2102153118. https://doi.org/10.1073/pnas.2102153118

Vazquez, F., Loscar, E. S., & Baglietto, G. (2019). Multistate voter model with imperfect copying. *Phys. Rev. E*, *100*, 042301. https://doi.org/10.1103/PhysRevE.100.042301

Verma, G., Swami, A., & Chan, K. (2014). The impact of competing zealots on opinion dynamics. *Physica A: Statistical Mechanics and its Applications*, *395*, 310–331. https://doi.org/https://doi.org/10.1016/j.physa.2013.09.045

Vicario, M. D., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2016). Echo chambers: Emotional contagion and group polarization on facebook. *Scientific Reports*, *6*. https://api.semanticscholar.org/CorpusID:8952182

Watts, D., & Dodds, P. (2007). Influentials, networks, and public opinion formation. *Journal of Consumer Research*, *34*, 441–458. https://doi.org/10.1086/518527

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world9 networks [Cited by: 34403]. *Nature*, *393*(6684), 440–442. https://doi.org/10.1038/30918

Weisbuch, G., Deffuant, G., & Amblard, F. (2004, October). *Persuasion dynamics*. https://doi.org/10.48550/arXiv.cond-mat/0410200

Weisbuch, G., Deffuant, G., Amblard, F., & Nadal, J.-P. (2002). Meet, discuss, and segregate! *Complexity*, *7*(3), 55–63. https://doi.org/https://doi.org/10.1002/cplx.10031

Williams, C. (2017). Introduction: Social media, political marketing and the 2016 u.s. election. *Journal of Political Marketing*. https://doi.org/10.1080/15377857.2017.1345828

Wong, Q. (2020). *Bloomberg's meme campaign underscores the loopholes in political-ad rules* [Accessed: 2024-11-11]. https://www.cnet.com/news/politics/bloombergs-meme-campaign-underscores-the-loopholesin-political-ad-rules/

Xia, H., Wang, H., & Xuan, Z. (2011). Opinion dynamics: A multidisciplinary review and perspective on future research. *Int. J. Knowl. Syst. Sci.*, *2*(4), 72–91. https://doi.org/10.4018/jkss.2011100106

Yuan, Y., Yan, J., & Zhang, P. (2021). Assortativity measures for weighted and directed networks. https://arxiv.org/abs/2101.05389