How opinions crystallise: an analysis of polarisation in the voter model

Antoine Vendeville, Benjamin Guedj, Shi Zhou

How opinions crystallise: an analysis of polarisation in the voter model

Antoine Vendeville
Centre for Doctoral Training in Cybersecurity
Department of Computer Science
University College London
London, United Kingdom
a.vendeville@ucl.ac.uk

Benjamin Guedj
Centre for Artificial Intelligence
Department of Computer Science
Inria & University College London
London, United Kingdom
b.guedj@ucl.ac.uk

Shi Zhou
Centre for Artificial Intelligence
Department of Computer Science
University College London
London, United Kingdom
s.zhou@ucl.ac.uk

Abstract

We address the phenomenon of sedimentation of opinions in networks. We investigate how agents who never change their minds ("stubborn") can influence the opinion of a social group and foster the formation of polarised communities. We study the voter model in which users are divided in two camps and repeatedly update their opinions based on others they connect with. Assuming a proportion of the agents are stubborn, the distribution of opinions reaches an equilibrium. We give novel formulas based on Markov Chain analysis to compute the distribution of opinions at any time and speed of convergence to stationary equilibrium. Theoretical results are supported by numerical experiments on synthetic data, and we discuss a strategy to mitigate the polarisation phenomenon.

1 Introduction

We are interested in how opinions evolve within a connected group of users. Namely, we study the well-known voter model, in which each user holds one of two possible opinions and updates it randomly under the distribution of others' beliefs. Independently introduced by Clifford and Sudbury (1973) and Holley and Liggett (1975) in the context of particles interaction, this model allows to describe in a simple and intuitive manner social dynamics where people are divided between two parties and form their opinion by observing that of others around them.

Online social platforms have gained a tremendous importance over the past decade. We use them to get in touch with each other and instantly share any type of content over the world. Combined with ease of access and use, this wide range of possibilities has made them leading actors of today's world. This is why getting a thorough understanding of how people connect and interact within social groups is becoming even more pressing and strategic.

We consider a group of users where everyone is influenced by everyone else, and assume some of them are stubborn and never change opinion. They can represent lobbyists, politicians or activists for example. Long time dynamics and expected limiting behaviour of such processes have been subject to several studies (Mobilia et al., 2007; Yildiz et al., 2013; Mukhopadhyay et al., 2020). Others (Pariser, 2011, amongst others) highlight in particular that this endeavours polarisation of opinions.

Preprint. Under review.

The novelty of our work is that we unveil step-by-step macrodynamics of the process through Markov Chain modelling and provide closed form formulas for the computation of convergence times.

In addition, we apply our findings to study the problem of reducing the polarisation of opinions. In recent years, recommendation algorithms on social platforms have greatly enhanced confirmation bias by showing users content that is the most susceptible to match their interests – the so-called "filter bubble" effect (Pariser, 2011). As a consequence more and more isolated, tightly clustered online communities of similar-minded individuals have arisen in various domains such as politics (Conover et al., 2011; Baer, 2016), science (McCright and Dunlap, 2011) or healthcare (Holone, 2016). We show that within the voter model framework, this phenomenon of opinions polarisation can be mitigated by adding links between fully-connected communities.

2 The voter model

Here we describe the classical voter model and we introduce the notation that will be used throughout the paper. We consider a group of n users labelled $1, \ldots, n$ who hold individual opinions in $\{0, 1\}$. These are prone to change over time and we let $x_i(t)$ denote the opinion of user i at time t. Each user has access to the opinion of some of the others, called her neighbours. Users can then be seen as forming a graph of n nodes, with an edge from j from i if and only if i has access to the opinion of j. The process unfolds as follows. Starting with a given initial distribution of opinions, an independent exponential clock of parameter 1 is associated to each user. Whenever a clock rings, the concerned user changes her opinion to that of one of her neighbours selected uniformly at random – or equivalently, chooses her new opinion by sampling the distribution of her neighbours' opinions. We say that consensus is reached if almost surely all users eventually hold the same opinion, *i.e.* for all $i, j \in \{1, \ldots, n\}$:

$$\mathbb{P}\left(x_i(t) = x_j(t)\right) \underset{t \to \infty}{\longrightarrow} 1.$$

We then have the following fundamental result.

Theorem 1. *Consensus is reached over any finite strongly connected social graph.*

We refer the reader to Aldous and Fill (2002, Chapter 14) for a proof based on coalescing random walks and Yildiz et al. (2010, Theorem 2.2) for another one using Markov Chain modelling. The intuitive idea is that no matter the current number of opinion-1 and opinion-0 holders, there exists a succession of individual opinion changes with strictly positive probability that results in everyone having the same opinion.

In this paper we place ourselves in clique, that is a community of users who all know each other. We will call it user or social graph without distinction. This results in the following assumption that holds throughout the paper, with the exception of Section 6.

Assumption 1. *The user graph we consider is a clique with unweighted edges and no self-loops.*

Thus each user only accounts for the opinion of everyone except their own with no particular preference. We look at macro dynamics of the voter model, that is, we consider the evolution of the size of each party (opinion 0 and opinion 1). To this end we let $N_1(t)$ denote the number of opinion-1 holders at time t; it will be our quantity of interest. We assume $N_1(0)$ is fixed and let n_1 denote its value. We are interested in the particular situation where some of the agents are stubborn, that is never change their opinions, and we describe the evolution of $N_1(t)$ over time. We denote by s_1 and s_0 the numbers of stubborn opinion-1 and opinion-0 agents respectively. We write $[m_{ij}]_{i,j}$ to denote the matrix with entry m_{ij} in the i-th row and j-th column and let e^M denote the exponential of any matrix M. Finally if Ω is a set we use $\mathcal{P}(\Omega)$ to indicate its powerset.

3 Related literature

Researchers have been studying the question of opinions evolution in social networks for almost fifty years. Perhaps the earliest well-known work in this area is from DeGroot (1974) who studied how a society of individuals may or may not come to an agreement on some given topic. Assuming the society is connected and people repeatedly update their belief by taking weighted averages of those of their neighbours, he showed that consensus is reached. Various other models have been developed since, to tackle the question of under which circumstances and how fast a population is able to reach consensus. Amongst others, Friedkin and Johnsen (1990) introduce immutable innate preferences, Axelrod (1997) studies the effect of confirmation bias, Banerjee and Fudenberg (2004) assume individuals are perfectly rational and Jadbabaie et al. (2012) account for the influence of external events.

The voter model that we consider here was introduced independently by Clifford and Sudbury (1973) and Holley and Liggett (1975) in the context of particles interaction. They proved that consensus is reached on the infinite $\mathbb{Z}^d$ lattice. Several works have since looked at different network topologies, wondering whether consensus is reached, on which opinion and at what speed. Complete graphs (Hassin and Peleg, 2002; Sood et al., 2008; Perron et al., 2009; Yildiz et al., 2010), Erdös-Rényi random graphs (Sood et al., 2008; Yildiz et al., 2010), scale-free random graphs (Sood et al., 2008; Fernley and Ortgiese, 2019), and other various structures (Yildiz et al., 2010; Sood et al., 2008) have been addressed. Variants where nodes deterministically update to the most common opinion amongst their neighbours have also been studied (Chen and Redner, 2005; Mossel et al., 2014).

An interesting case to consider is the one where stubborn agents who always keep the same opinion are present in the graph. Such agents may for example represent lobbyists, politicians or activists, *i.e.* entities looking to lead rather than follow and who will not easily change side. One of those placed within the network can singlehandedly change the outcome of the process (Mobilia, 2003; Sood et al., 2008). If several of them are present on both sides, consensus is usually not reachable and instead opinions converge to a steady-state in which they fluctuate indefinitely (Mobilia et al., 2007; Yildiz et al., 2013).

Recently, Mukhopadhyay et al. (2020) considered agents with different degrees of stubbornness and show that time to reach consensus grows linearly with their number. They also show that if one opinion is initially preferred – *i.e.* agents holding that opinion have a lesser probability of changing their mind – consensus is reached on the preferred opinion with a probability that converges to 1 as the network size increases. Klamser et al. (2017) study the effect of stubborn agents on a dynamically evolving graph, and show that the two main factors shaping their influence are their degrees and the dynamical rewiring probabilities.

Lastly, theoretical solutions have been recently proposed to minimise polarisation of opinions while preserving social structures as much as possible. Yi and Patterson (2019) formulate different constrained optimisation problems under the Degroot and the Friedkin-Johnsen models. They provide solutions in the form of optimal graph construction methods. Chitra and Musco (2020) consider the Friedkin-Johnsen model and study the problem from the network administrator perspective. They prove that dynamically nudging edge weights in the user graph can reduce polarisation while preserving relevance of the content shown by the recommendation algorithm.

Our contributions. We study the impact of stubborn agents in the voter model when users form a complete graph. We perform a novel Markovian analysis and develop closed form formulas for the computation of exact convergence times and evolution of opinions distribution. To the best of our knowledge this is the first time such work is conducted. We also apply our findings to provide empirical evidence that adding random links between fully-connected communities can mitigate polarisation. All code used for the simulations is available at `https://github.com/AntoineVendeville/HowOpinionsCrystallise`.

Outline. In Section 4 we assume stubborn agents are found on one side only. In Section 5 we extend our analysis to the case where each camp counts a few of them. Simulation results are shown to back up the theory. Section 6 is devoted to our take on mitigating polarisation. In Section 7 we conclude and discuss leads for future work.

4 Stubborn agents in one camp

4.1 Modelling

We consider the voter model as presented in Section 2 with a complete graph of n users amongst whom $n_1 > 0$ initially hold opinion-1. Additionally in this section, we assume that some of them are stubborn and will never change their minds. They form an inflexible core of partisans with great power of persuasion over the whole population. This is formalized through Assumption 2.

Assumption 2. *A number s_1 of opinion-1 holders are stubborn. The remaining $n - s_1$ users are free to change their opinions.*

Consequently the process unfolds as follows. We start by choosing uniformly at random n_1 users to hold opinion 1 and s_1 amongst them to be stubborn. Each of the n users is associated with an exponential clock of parameter 1. Whenever a clock rings, if the associated user is not stubborn she chooses another user uniformly at random and switches her opinion to theirs. Equivalently, at the times of a Poisson process of parameter n, we select a user uniformly at random and if she is not stubborn, she changes her opinion to that of another user drawn uniformly at random.

Let $N_1(t)$ denote the number of users with opinion 1 at time t. The main idea at the basis of our analysis is that it describes a birth-and-death process over the state-space $\{s_1, \dots, n\}$ with transition rates

$$\begin{cases} q_{k,k-1} = (k - s_1)(n-k)/(n-1) \\ q_{k,k+1} = k(n-k)/(n-1) \\ q_{k,k} = -q_{k,k-1} - q_{k,k+1} \end{cases} \tag{1}$$

for all $s_1 \leqslant k \leqslant n-1$. Indeed to move from state k to $k-1$ we need a non stubborn opinion-1 holder to adopt the opinion of an opinion-0 user. There are $k - s_1$ non stubborn opinion-1 users and a proportion $(n-k)/(n-1)$ of opinion-0 users, hence $q_{k,k-1} = (k-s_1)(n-k)/(n-1)$. We obtain $q_{k,k+1}$ via an analogous reasoning. Since the process only evolves by unit increments or decrements, $q_{k,j} = 0$ if $j \notin \{k-1, k, k+1\}$. Moreover we have $q_{n,j} = 0$ for all j and let $Q = [q_{ij}]_{i,j}$ denote the transition rate matrix. Finally since n is the only absorbing state and is reachable from any other state, we have the following consensus guarantee.

Theorem 2. *If Assumption 2 holds then almost surely all users end up with opinion 1 in finite time.*

This is not surprising. Indeed we know (Theorem 1) that consensus is reached over any finite graph without stubborn agents. Now if some users are not willing to abandon opinion 1, consensus cannot be reached on opinion 0 and thus opinion 1 eventually prevails.

4.2 Expectation and convergence time

Two natural questions that come to mind are what is the expected value of $N_1(t)$ at any time, and how long does it take for consensus to be reached. We provide answers to both. From Norris (1997, Theorems 2.1.1 and 2.8.2) we know that the probability for a continuous-time Markov Chain with transition rate matrix Q to go from state k to state ℓ in a time interval of length t is given by the (k, ℓ)-th entry of the matrix exponential e^{tQ}. Hence the expectation of $N_1(t)$ at any time can be computed as follows.

Theorem 3. *Let Q be the matrix with entries described in (1) and let $N_1(0) = n_1$ be given. Assuming there are $s_1 > 0$ stubborn agents with opinion-1, the expected number of opinion-1 holders at time t is given by*

$$\mathbb{E}N_1(t) = \sum_{k=s_1}^{n} k \, p_{n_1,k}(t). \tag{2}$$

where $p_{n_1,k}(t) := [e^{tQ}]_{n_1,k}$ is the probability for the number $N_1(t)$ of opinion-1 holders to equal k at time t.

Note that this is straightforwarldy obtained by combining results from Norris (1997, Theorems 2.1.1 and 2.8.2). Regarding convergence time, we want to find the expected time before state $N_1(t)$ is in state n, from where it will not be able to move anymore. This is called the expected hitting time of state n, and it can be computed using (Norris, 1997, Theorem 3.3.3).

Theorem 4. *Let Q be the matrix with entries described in (1). Let $h_k^{(n)}$ denote the expected hitting time of state n when starting from state k. Then $(h_{s_1}^{(n)}, \dots, h_n^{(n)})$ is solution of the linear system*

$$\begin{cases} h_n^{(n)} = 0, \\ \sum_{j=s_1}^{n-1} q_{i,j} h_j^{(n)} = -1, \quad \text{for } i = s_1, \dots, n-1. \end{cases} \tag{3}$$

Remark that because Q is tridiagonal, only three values in each sum are non zero. Thus the expected hitting time from any state depends only on those from neighbouring states and transition rates to these state.

4.3 Simulations

We now provide empirical evidence and visual support for Theorem 3 and Theorem 4 through numerical experiments. We place ourselves in a complete network with $n = 1000$ users. We start with $n_1 = 250$ opinion-1 holder amongst whom there is $s_1 = 1$ stubborn agent. We let the process evolve for 150 time units. We repeat over $m = 100$ simulations and then do the same for $s_1 = 10$ and $s_1 = 250$. In Figure 1 **(left)** we plot the empirical number of opinion-1 holders at each step averaged over all simulations with corresponding theoretical values obtained through Equation 2. Values were calculated every 10 time units. We also show confidence intervals of the form

$$\left[\overline{N_1(t)} - \phi\frac{\hat{\sigma}(N_1(t))}{\sqrt{m}},\ \overline{N_1(t)} + \phi\frac{\hat{\sigma}(N_1(t))}{\sqrt{m}}\right] \tag{4}$$

where $\overline{N_1(t)}$ and $\hat{\sigma}(N_1(t))$ are respectively the empirical mean and the unbiased empirical standard deviation of $N_1(t)$ over all simulations, m is the number of steps and ϕ is the 95% quantile of the normal distribution. First of all, we notice that experiments supports the analysis as theoretical values fall within the range of confidence intervals constructed from empirical means. Moreover we see that the number of stubborn agents has an important impact on the dynamics of the process. With only one of them $N_1(t)$ barely increases by 50% over 150 time units while when they are 250, consensus is reached very early – roughly 30 times units, which corresponds to a 400% increase in $N_1(t)$.

In Figure 1 **(right)** we plot expected hitting times of state n function of the initial number of opinion-1 holders. For n_1 in $\{250, 600, 900\}$ we show theoretical values obtained via Theorem 4, alongside empirical averages over $m = 100$ simulations with confidence bars from Equation 4. Here again theoretical values fall within the range of confidence intervals constructed from empirical means. We observe that the difference in convergence speed between different values of s_1 is more sensible as n_1 get smaller. When only 10 of the users initially hold opinion-1, them all being stubborn compared to just one will roughly reduce time to consensus by a factor of 5.

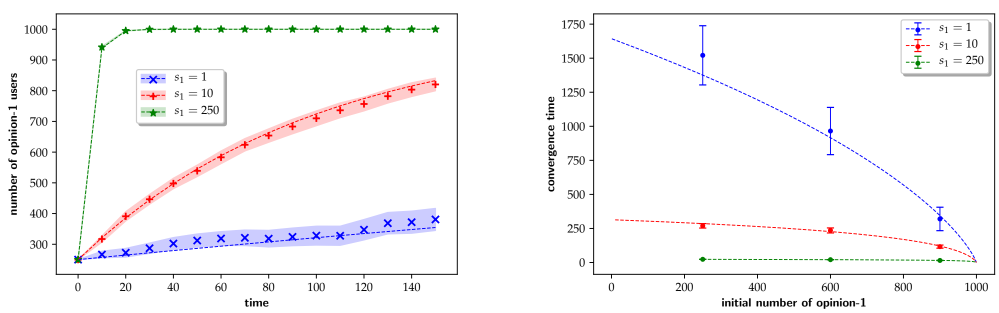

Figure 1: Complete graph with 1000 users. **Left:** markers represent the empirical number of opinion-1 holders at each step averaged over 100 simulations, while dotted lines follow theoretical values from Equation 2. Shaded areas are confidence intervals obtained via Equation 4. **Right:** convergence speed function of n_1. Bars represent empirical values averaged over 100 simulations with confidence intervals, while dotted lines follow theoretical values from Theorem 4.

5 Stubborn agents in both camps

5.1 Modelling

We extend the previous analysis by injecting $s_0 > 0$ stubborn agents in the opinion-0 camp, in addition to $s_1 > 0$ stubborn ones with opinion 1. The remaining $n - s_0 - s_1$ users are free to change opinions. In that case consensus cannot be reached, as there will always be a stricly positive number of users supporting each opinion. Rather, opinions reach a state of equilibrium in which they fluctuate indefinitely (Mobilia et al., 2007; Yildiz et al., 2013). The number $N_1(t)$ of opinion-1 holders at time t describes a birth-and-death process over the state-space $\{s_1, \dots, n - s_0\}$ with no absorbing state and the following transition rates for all $s_1 \leqslant k \leqslant n - s_0$:

$$\begin{cases} q_{k,k-1} = (k - s_1)(n - k)/(n - 1) \\ q_{k,k+1} = k(n - k - s_0)/(n - 1) \\ q_{k,k} = -q_{k,k-1} - q_{k,k+1}. \end{cases} \tag{5}$$

Values are obtained via a reasoning analogous as in the previous section. We have $q_{k,j} = 0$ if $j \notin \{k-1, k, k+1\}$ and $q_{n,j} = 0$ for all j.

5.2 Expectation and steady-state

As in the previous section we aim at finding the expectation of $N_1(t)$ and evaluate time of convergence to equilibrium. We start by looking at the first. To this end we adapt Theorem 3 with minor updates.

Theorem 5. *Let Q be the matrix with entries described in (5) and let $N_1(0) = n_1$ be given. Assuming there are $s_0, s_1 > 0$ stubborn agents with respective opinions 0 and 1, the expected number of opinion-1 holders at time t is given by*

$$\mathbb{E}N_1(t) = \sum_{k=s_1}^{n-s_0} k\, p_{n_1,k}(t). \tag{6}$$

where $p_{n_1,k}(t) := [e^{tQ}]_{n_1,k}$ is the probability for the number $N_1(t)$ of opinion-1 holders to equal k at time t.

This theorem can also be used to characterise the equilibrium of the process, but there exist more interpretable ways. It is known that the stationary distribution π of a Markov Chain with transition rates matrix Q is solution of the linear system $\{\pi Q = 0, \|\pi\|_1 = 1\}$, and even simpler formulas exist for birth-and-death processes (Bocharov et al., 2011, section 1.5.7). Regarding the expectation at equilibrium, we prove the following.

Theorem 6. *Assuming there are $s_0, s_1 > 0$ stubborn agents with respective opinions 0 and 1, the expected number of opinion-1 holders at equilibrium is given by*

$$\mathbb{E}\pi = n \frac{s_1}{s_0 + s_1} \tag{7}$$

where $\pi = (\pi_{s_1}, \ldots, \pi_{n-s_0})$ denotes the steady-state distribution of $N_1(t)$.

Proof. We know (Yildiz et al., 2013, Theorem 2.1) that the vector of individual opinions $(x_1(t), \ldots, x_n(t))$ converges in distribution to a random vector $(x_1^*, \ldots, x_n^*)$. From Yildiz et al. (2013, Proposition 3.2) we have that $\mathbb{E}x_i^*$ is equal to the probability that a random walk on the user graph initiated at node i is absorbed by the set of stubborn opinion-1 agents. Here we are in a complete graph where all nodes are topologically equivalent. Thus the probability that a random walk on the user graph starting at node i hits a stubborn opinion-1 agent before an opinion-0 one is simply $s_1/(s_0 + s_1)$. Hence $\mathbb{E}x_i^* = s_1/(s_0 + s_1)$. Since there are $n - s_0 - s_1$ such non-stubborn nodes and s_1 stubborn ones with opinion 1, we have

$$\mathbb{E}\pi = s1 + (n - s_0 - s_1)\, \mathbb{E}x_i^* \tag{8}$$

and Theorem 6 ensues. □

This theorem states that the proportion of opinion-1 users is expected to endlessly fluctuate around the ratio $s_1/(s_0 + s_1)$. For example, having twice as many stubborn agents as the other camp will on average lead to count twice as many partisans. Thus the camp that boasts the biggest quantity of stubborn agents is expected to be of bigger size in the long run.

5.3 Convergence time

Given that there is no absorbing state here, the question of convergence speed requires different gear that the expected hitting times we used previously. We look at the mixing time of our Markov Chain, defined by

$$T_\varepsilon(n_1) = \inf \left\{ t > 0,\, \|p_{n_1,\cdot}(t) - \pi\| < \varepsilon \right\}. \tag{9}$$

Intuitively speaking, this represents the shortest time length for the chain to be at distance lower than ε from its stationary distribution. An abundant literature suggests to use the distance in total variation, which for any two probability measures μ, ν on a set Ω is defined as

$$\|\mu - \nu\|_{TV} = \max_{\mathcal{A} \in \mathcal{P}(\Omega)} |\mu(\mathcal{A}) - \nu(\mathcal{A})|. \tag{10}$$

In other words, total variation measures the biggest possible "disagreement" between μ and ν over all events. As the set $\mathcal{P}(\Omega)$ is of size $2^{|\Omega|}$, computing the above is intractable in most cases. Thus we use a more convenient characterisation (Levin et al., 2017, Proposition 4.2):

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|. \tag{11}$$

This formula requires only $|\Omega|$ computations, making it much more usable in practice. The convergence speed of the model is then obtained via the following.

Theorem 7. *The mixing time of $N_1(t)$ with precision ε starting from state n_1 is given by*

$$T_\varepsilon(n_1) = \inf \left\{ t > 0, \sum_{k=s_1}^{n-s_0} |\, p_{n_1,k}(t) - \pi_k| < 2\varepsilon \right\}. \tag{12}$$

We now validate and illustrate Theorems 5, 6, 7 through numerical experiments.

5.4 Simulation

We place ourselves in a complete network with $n = 1000$ users. For $(s_0, s_1) = (100, 50)$ and $(200, 250)$ we run $m = 100$ simulations with starting values $n_1 = 250, 750$. Results are shown in Figure 2 **(left)**. We plot the empirical proportion of opinion-1 holders at each step averaged over all simulations with confidence intervals (4) and corresponding theoretical values from Equation 6. Values were calculated every 2 time units. We also show the expected limiting number of opininion-1 holders obtained via Theorem 6. Experiments supports the analysis as theoretical values fall within the range of confidence intervals constructed from empirical means, and $N_1(t)$ converges towards its limiting expectation $ns_1/(s_0 + s_1)$ regardless of the initial state n_1. Also, equilibrium is reached quicker with $(s_0, s_1) = (200, 250)$. A higher number of stubborn agents leaves less room for overall change, making the system closer to stability.

Additionally in Figure 2 **(right)** we show theoretical mixing times function of n_1. We take $(s_0, s_1) = (100, 50)$ then $(200, 250)$ and let the precision threshold ε span $\{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$. The sum in Equation 12 was evaluated every $dt = 0.1$ time unit starting from $t = 0$, yielding an approximation of $T_\varepsilon(n_1)$ with error margin dt. Mixing times look symmetrical around the limiting expectation, where they drop dramatically. Also, note that the constant gaps between curves might be due to the regular spacing we chose for ε.

6 Reducing polarisation

We apply our model to study the problem of reducing polarisation. Namely, we look at two strongly-opinionated, isolated groups of users and wonder how the addition of links between them can moderate the balance of opinions. Consider a social graph divided in two cliques C_0 and C_1, *i.e.* two internally fully connected groups of users, with no links between them. Each clique supports a different opinion, users in C_0 being more inclined towards opinion 0 and those in C_1 towards 1. To account for this, C_0 counts a few stubborn agents with opinion 1, none with opinion 0 and an initial majority of opinion-1 holders. Simarly for C_1 with opposite opinions. Were there no links between the two, according to Theorem 2 everyone in C_0 would eventually hold opinion 0 and everyone in C_1 opinion 1. We now show that adding connections between the groups can mitigate this effect, as we conduct simulations of the voter model on the whole graph with varying numbers of links between C_0 and C_1.

For our experiments we consider two cliques of 500 people each, thus forming a graph with 1000 total users. We first look at a symmetric case: C_0 has 400 initial opinion-0 holders amongst whom 250 are stubborn, while clique C_1 has 400 initial opinion-1 holders amongst whom 250 are stubborn.

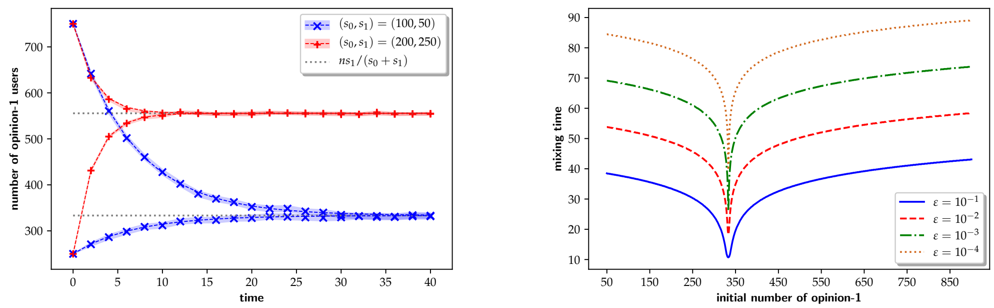

Figure 2: Complete graph with 1000 nodes. **Left:** for each of $n_1 = 250$ and $n_1 = 750$ we take $(s_0, s_1) = (100, 50)$ and $(200, 250)$. Markers represent the empirical proportion of opinion-1 holders at each step averaged over 100 simulations, while dashed colored lines stand for corresponding theoretical values from Equation 6. Shaded areas are confidence intervals obtained via Equation 4. Expected limiting values from Theorem 6 are plotted as constant grey dashed lines. **Right:** Theoretical mixing time function of n_1 with $(s_0, s_1) = (100, 50)$, obtained via Theorem 7.

We add 12,500 links at random between the cliques (5% of all possible ones) and simulate the voter model over 20 times unit. We do the same with 15% and finally 75% of all possible links. In Figure 3 **(left)** we plot the number of opinion-1 holders over time for each clique, averaged over $m = 100$ simulations and with confidence intervals (Equation 4). We also show theoretical values of $\mathbb{E}N_1(t)$ were there no in-between links at all (obtained via Equation 2).

What we see is that adding links between the cliques indeed mitigates polarisation. A mere 5% of all possible ones already yields a visible effect and prevents consensus to the favoured opinion in each clique. Preferences are preserved, as putting even 75% of all possible links between the cliques still leaves them strongly-opinionated. In Figure 3 **(right)** we break the symmetry by reducing the number of stubborn opinion-1 agents in clique C_1 to 50. In this case there is a clear unbalance between the cliques. Users in C_1 are less resistant to the attraction of opinion 1, and drawing 75% of all possible in-between links is enough to make the clique switch sides and globally favour opinion 0. On the other hand, while consensus is prevented, polarisation in C_0 remains very strong.

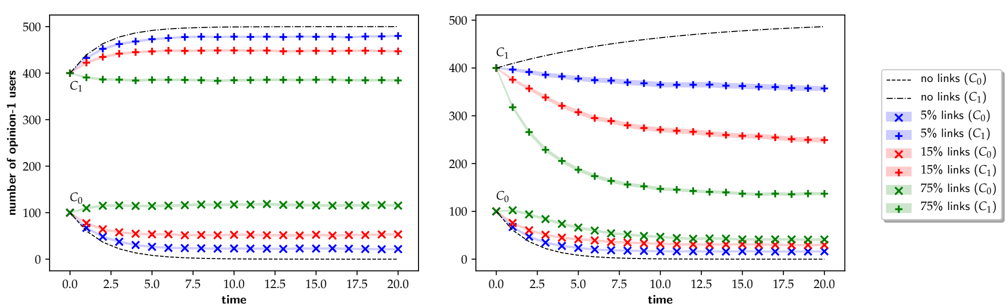

Figure 3: Polarisation reduction. Cliques C_0 and C_1 both count 500 users, the first favours opinion 0 and the second opinion 1. We simulate with different numbers of links between the cliques. Markers represent empirical averages over 100 simulations for the number of opinion-1 holders in each clique, taken every time unit. Shaded areas delimit confidence interval (Equation 4). Dotted lines represent theoretical values when there are no links between the cliques. **Left:** C_0 has 400 initial opinion-0 holders amongst whom 250 are stubborn, clique C_1 has 400 initial opinion-1 holders amongst whom 250 are stubborn. **Right:** Identical setup except C_1 has only 50 stubborn opinion-1 agents.

7 Conclusion and future work

We have analysed the voter model on complete graphs with stubborn agents. We have given closed form formulas for the expected number of opinion-1 holders over time and for convergence times. In the case were stubborn agents are present on both sides we have also characterised the stationary equilibrium of the process. Our analysis was supported through numerical simulations. Additionally we have shown that polarisation of opinions can be reduced by adding links between strongly-opinionated communities. Leads for further work include extending our analysis to other graph models, considering different degrees of stubbornness and confirming the validity of our approach through applications to real-life datasets.

Acknowledgments and Disclosure of Funding

This project was funded by the UK EPSRC grant EP/S022503/1 that supports the Centre for Doctoral Training in Cybersecurity delivered by UCL's Departments of Computer Science, Security and Crime Science, and Science, Technology, Engineering and Public Policy.

References

D. Aldous and J. A. Fill. Reversible markov chains and random walks on graphs, 2002. Unfinished monograph, recompiled 2014, available at http://www.stat.berkeley.edu/~aldous/RWG/book.html.

R. Axelrod. The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution*, 41(2):203–226, 1997. doi: 10.1177/0022002797041002001. URL https://doi.org/10.1177/0022002797041002001.

D. Baer. The 'Filter Bubble' explains why Trump won and you didn't see it coming. *Science Of Us*, 2016.

A. Banerjee and D. Fudenberg. Word-of-mouth learning. *Games and Economic Behavior*, 46 (1):1–22, Jan. 2004. ISSN 08998256. doi: 10.1016/S0899-8256(03)00048-4. URL https://www.sciencedirect.com/science/article/pii/S0899825603000484.

P. P. Bocharov, C. D'Apice, and A. V. Pechinkin. *Queueing Theory*. De Gruyter, 2011. ISBN 978-3-11-093602-5. URL https://www.degruyter.com/view/title/19480.

P. Chen and S. Redner. Majority rule dynamics in finite dimensions. *Phys. Rev. E*, 71:036101, Mar 2005. doi: 10.1103/PhysRevE.71.036101. URL https://link.aps.org/doi/10.1103/PhysRevE.71.036101.

U. Chitra and C. Musco. Analyzing the impact of filter bubbles on social network polarization. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, WSDM '20, page 115–123, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450368223. doi: 10.1145/3336191.3371825. URL https://doi.org/10.1145/3336191.3371825.

P. Clifford and A. Sudbury. A model for spatial conflict. *Biometrika*, 60(3):581–588, 12 1973. ISSN 0006-3444. doi: 10.1093/biomet/60.3.581. URL https://doi.org/10.1093/biomet/60.3.581.

M. Conover, J. Ratkiewicz, M. Francisco, B. Goncalves, F. Menczer, and A. Flammini. Political polarization on twitter. In *International AAAI Conference on Web and Social Media*, 2011. URL https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2847.

M. H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345): 118–121, 1974. ISSN 01621459. URL http://www.jstor.org/stable/2285509.

J. Fernley and M. Ortgiese. Voter models on subcritical inhomogeneous random graphs, 2019. URL https://arxiv.org/abs/1911.13187.

N. E. Friedkin and E. C. Johnsen. Social influence and opinions. *The Journal of Mathematical Sociology*, 15(3-4):193–206, Jan. 1990. ISSN 0022-250X, 1545-5874. doi: 10.1080/0022250X.1990.9990069. URL http://www.tandfonline.com/doi/abs/10.1080/0022250X.1990.9990069.

Y. Hassin and D. Peleg. Distributed probabilistic polling and applications to proportionate agreement. *Inf. Comput.*, 171(2):248–268, Jan. 2002. ISSN 0890-5401. doi: 10.1006/inco.2001.3088. URL https://doi.org/10.1006/inco.2001.3088.

R. A. Holley and T. M. Liggett. Ergodic theorems for weakly interacting infinite systems and the voter model. *Ann. Probab.*, 3(4):643–663, 08 1975. doi: 10.1214/aop/1176996306. URL https://doi.org/10.1214/aop/1176996306.

H. Holone. The filter bubble and its effect on online personal health information. *Croatian medical journal*, 2016. doi: 10.3325/cmj.2016.57.298. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4937233/.

A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi. Non-bayesian social learning. *Games and Economic Behavior*, 76(1):210 – 225, 2012. ISSN 0899-8256. doi: https://doi.org/10.1016/j.geb.2012.06.001. URL http://www.sciencedirect.com/science/article/pii/S0899825612000851.

P. P. Klamser, M. Wiedermann, J. F. Donges, and R. V. Donner. Zealotry effects on opinion dynamics in the adaptive voter model. *Phys. Rev. E*, 96:052315, Nov 2017. doi: 10.1103/PhysRevE.96.052315. URL https://link.aps.org/doi/10.1103/PhysRevE.96.052315.

D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times, second edition*. American Mathematical Society, 2017.

A. M. McCright and R. E. Dunlap. The politicization of climate change and polarization in the american public's views of global warming, 2001–2010. *The Sociological Quarterly*, 52(2):155–194, 2011. doi: 10.1111/j.1533-8525.2011.01198.x. URL https://doi.org/10.1111/j.1533-8525.2011.01198.x.

M. Mobilia. Does a single zealot affect an infinite group of voters? *Phys. Rev. Lett.*, 91:028701, Jul 2003. doi: 10.1103/PhysRevLett.91.028701. URL https://link.aps.org/doi/10.1103/PhysRevLett.91.028701.

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

E. Mossel, J. Neeman, and O. Tamuz. Majority dynamics and aggregation of information in social networks. *Autonomous Agents and Multi-Agent Systems*, 28(3):408–429, May 2014. ISSN 1387-2532. doi: 10.1007/s10458-013-9230-4. URL https://doi.org/10.1007/s10458-013-9230-4.

A. Mukhopadhyay, R. R. Mazumdar, and R. Roy. Voter and majority dynamics with biased and stubborn agents, 2020. URL https://arxiv.org/abs/2003.02885.

J. R. Norris. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997. doi: 10.1017/CBO9780511810633.

E. Pariser. *The Filter Bubble: What the Internet Is Hiding from You*. The Penguin Group, 2011. ISBN 1594203008.

E. Perron, D. Vasudevan, and M. Vojnovic. Using three states for binary consensus on complete graphs. In *IEEE INFOCOM 2009*, pages 2527–2535, 2009. URL https://ieeexplore.ieee.org/document/5062181.

V. Sood, T. Antal, and S. Redner. Voter models on heterogeneous networks. *Physical Review E*, 77 (4), Apr 2008. ISSN 1550-2376. doi: 10.1103/physreve.77.041121. URL http://dx.doi.org/10.1103/PhysRevE.77.041121.

Y. Yi and S. Patterson. Disagreement and polarization in two-party social networks, 2019. URL https://arxiv.org/abs/1911.11338.

E. Yildiz, A. Ozdaglar, D. Acemoglu, A. Saberi, and A. Scaglione. Binary opinion dynamics with stubborn agents. *ACM Trans. Econ. Comput.*, 1(4), Dec. 2013. ISSN 2167-8375. doi: 10.1145/2538508. URL https://doi.org/10.1145/2538508.

M. E. Yildiz, R. Pagliari, A. Ozdaglar, and A. Scaglione. Voting models in random networks. In *2010 Information Theory and Applications Workshop (ITA)*, pages 1–7, Jan 2010. doi: 10.1109/ITA.2010.5454090. URL https://ieeexplore.ieee.org/document/5454090.