UNIVERSITY OF CALIFORNIA,
IRVINE

Mathematical Theory of Opinion Dynamics with Applications

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Mathematics

by

Daniel Vincent Simonson

Dissertation Committee:
Professor Natalia Komarova, Chair
Professor German Enciso
Professor John Lowengrub

2022

DEDICATION

To Gabriela, without whom I would not be where I am today.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank the Department of Mathematics at University of California, Irvine, for their support, without which none of this would have been possible. I also thank my daughter for her love and patience with me as I split time between being a father to her and the work involved that culminated in the present dissertation. Additionally, I thank Professor Chen Li and his research team for gathering and sharing with me the data set used in chapter 4. Furthermore, I wish to thank my friend Rachel, with whom I've had many conversations about this project, and for the different perspectives from which she helped me to consider in this work. Most of all I wish to thank my advisor, Professor Natalia Komarova for her on-going encouragement, advice, and kindness throughout this entire journey.

VITA

Daniel Vincent Simonson

EDUCATION

Doctor of Philosophy in Mathematics **2022**
University of California, Irvine *Irvine, California*

Master of Science in Mathematics **2017**
University of California, Irvine *Irvine, California*

Bachelor of Science in Statistics **2014**
Sonoma State University *Rohnert Park, California*

Bachelor of Arts in Mathematics **2014**
Sonoma State University *Rohnert Park, California*

TEACHING EXPERIENCE

Teaching Associate **2021**
University of California, Irvine *Irvine, California*

Teaching Assistant **2015–2021**
University of California, Irvine *Irvine, California*

ABSTRACT OF THE DISSERTATION

Mathematical Theory of Opinion Dynamics with Applications

By

Daniel Vincent Simonson

Doctor of Philosophy in Mathematics

University of California, Irvine, 2022

Professor Natalia Komarova, Chair

Opinion dynamics can be modeled by using agent-based simulations, where agents in a population are characterized by binary opinions on a number of different issues. They engage in pairwise interactions, whereby if the agreement level is high, the interlocutor is recognized as an "ally" and the individual will flip one of their opinions to coincide with the interlocutor; if the agreement is low, they will switch away from the interlocutor. While it is usually assumed that all issues in the opinion vector are equally important, in chapter 2 we investigate how breaking this symmetry influences the dynamics. We find that the model outcomes can be predicted by a single Agreement-Disagreement Score (ADS) in $[-1, 1]$. ADS characterizes how likely individuals in the population are to regard an interlocutor as an ally; low-ADS (very "cautious") populations tend to converge to a two-faction system with exponentially high convergence times, while high-ADS (very "trusting") populations tend to converge to a single-faction system relatively fast. In heterogeneous populations characterized by individual issue weighting, individuals that are more "trusting" are more likely to join the majority group compared to those that are more "cautious". In the presence of an influencer, for ADS both near -1 and 1, a single faction tends to emerge, but in the former case it coincides with the influencer's opinions, while in the latter case it is the opposite. Time to fixation is also affected by the presence of an influencer, especially for negative-ADS populations, where it no longer experiences such a large increase near

-1. One can say that an influencer unifies the population to align with the source of influence if ADS> 0 and to disagree with it if ADS< 0, and consensus is reached relatively fast for both extremely "trusting" and extremely "cautious" populations. In chapter 3 we introduce a system of ordinary differential equations (ODEs) which govern the behavior of the discrete stochastic model studied in chapter 2. We find a neutrally stable solution to the system which, when it is the only stable solution, provides a mathematical description for the extremely long times to fixation observed in the stochastic model under certain conditions. Chapter 4 turns to a real world example as we analyze the so-called re-tweet network of tweets concerning the HPV vaccine on the social network Twitter. Community detection algorithms provide a way to split the nodes of a network into two or more communities. Using a label propagation method, survey data, and a construction of cultural consensus for survey response data we find a natural division in this network into two distinct communities, one of which is pro-HPV vaccine and the other is anti-HPV vaccine.

Chapter 1

Introduction

In 1997, Robert Axelrod introduced his highly influential opinion dynamics model [1] in an effort to show, under a minimal set of assumptions, that pockets of differences can persist within an overall systemic tendency towards uniformity. Axelrod was interested in cultural diffusion, so in his model, "agents" were cultural units, such as villages, which were located on a square grid, and opinions on issues were expressions of cultural traits. Each village was represented as an F-dimensional vector where each of the F components was a cultural feature taking on an integer value in $\{1, \ldots, M\}$. The value of a cultural feature was called a trait. Under the assumptions that (1) similarities arise from interaction, and (2) interaction is more likely between agents who are relatively more similar, the state of the system was updated thus: at each discrete unit in time, a random village was chosen to be active, and one of its neighbors was chosen at random. With probability equal to the fraction of common features the active village decided to interact with the neighbor, whereby it chose a random feature whose value disagreed with the neighbor, and switched to agree with the neighbor. The model showed that more traits per feature resulted in larger numbers of distinct cultural regions, while trials with fewer traits per feature tended to result in consensus. Allowing for larger neighborhoods of interaction also resulted in a

tendency towards consensus.

Opinion dynamics is an active field of study that has been developing in several different directions, see e.g. an excellent review [25]. The wealth of models studied in this context can be roughly classified as those containing a one- vs several-dimensional opinion vector; those with discrete vs continuous opinion values; and those with or without an external information source. Here we focus on models with opinion vectors that contain more than one discrete-valued component. In Axelrod's original model and many of its extensions, the dynamics are characterized by the active site's tendency to interact with its neighbors, weighted by similarity. Under this assumption [11] showed that if $F = M = 2$, then the majority of the population forms one cultural cluster, and if $F < M$ then fragmentation of the population persists. [12] found parameters for fixation of the process in a unidimensional region. Here fixation is taken to mean that each site/agent updates a finite number of times before the system freezes and no further updates are possible. [2] performed a thorough analysis of the Axelrod model, and found a phase transition from a culturally polarized (ordered) state to one which was culturally fragmented (disordered) state, as well as a distribution of the sizes of cultural regions. In [7, 18] the dynamics were governed by agents who formed their opinions based on the averaging (of different types) of opinions of those agents who were not too far from them in the opinion space (that is, within a certain confidence bound). [8] analyzed the model in the non-spatial case of a complete network. The assumption that similar individuals are more likely to interact was modeled as an influence graph on the set of types. Individuals of similar types were connected by a relatively higher weight than dissimilar types. It was shown that all initial conditions lead to a fixation in which the population evolves to separate non-interacting groups. [19] introduced disagreement dynamics, in which a parameter $\alpha \in [0, 1]$ is fixed and attractive dynamics (as in the Axelrod's model) are used if similarity is greater than or equal to α. Otherwise repulsive dynamics are used where the active site chooses a random feature that is in agreement with the neighbor, and changes it. These simulations were run on a

2-dimensional space (a grid as in Axelrod's original model). Plotting the average size of the largest cultural region against the number of traits per feature revealed near vertical transitions at different points for different values of α.

Several groups studied the influence of external sources and other inhomogeneities of behavior on the opinion dynamics. [24] introduced "committed" individuals, whose opinion cannot be changed. This was done in conjunction with the assumption of co-evolving networks, which is a way to account for homophily, the tendency to connect with similar individuals. It was shown that adding committed individuals at a fraction above a threshold could lead to an increase in time to consensus. Several papers introduced access to external information where at each step, instead of interacting with peers, each agent could interact with an information source [6, 4, 16, 17]. A large probability to be exposed to an information source lead to a larger degree of fragmentation of the system, or a "negative publicity" effect, where the population became homogenized "against" the information source. More sophisticated ways to set up interactions between individuals and external information sources were also implemented [5]. In [21], the relevance of each source to an individual was quantified, representing in a stylized fashion the targeting of different groups by the media. This produced a number of regimes, one resulting in the population homogenization where the individuals were aligned with the information source. Cultural ordering was also found (in the case where noise was below a threshold) in [14]. In [22], the concept of "social influence" was investigated to study the effects of mass media; different regimes were observed, from a monoculture to a fragmented population, depending on the initial diversity level.

Throughout these studies, the weights of different components of an agent's opinion vector were always assumed identical. In reality, each individual holds opinions about a variety of issues, but not all of them are equally important. For example, when comparing one's opinion vector to another agent's opinion vector, it appears more realistic to weigh some

issues more than others. In chapter 2 we introduce an opinion dynamics model where the issues are not all weighted equally. We study whether or not consensus occurs, how large the largest group is, and how long fixation takes under different parameter choices. We compare the case in which the individuals all weigh issues the same with that in which the individuals weigh the issues differently; finally we study the effects of outside influence on this model. In chapter 3 we introduce a system of ODEs which are a deterministic analog of the model developed in chapter 2 and use this system to further explain some of the behavior observed in the model. Finally, chapter 4 analyzes the resulting network arising from opinion dynamics in interactions concerning the HPV vaccine on the social media platform Twitter.

Chapter 2

The effects of opinion weighting, (dis)agreement, and external influence on social group formation

2.1 Introduction

In the present chapter we introduce an agent-based model of opinion dynamics in which the issues are not all weighted equally. We allow for both agreement and disagreement dynamics, i.e. we allow for conditions in which agents in the community may either change opinions to agree or to disagree with an interlocuter. We define a parameter called the *Agreement-Disagreement Score* (ADS) which measures how likely a community member is to undergo agreement or disagreement dynamics and explore the extent to which ADS explains the resulting fixed state for communities in which all individuals weigh the issues the same versus communities where individuals weigh issues differently; communities in which there is no outside influence versus communities with an external influence; and small communities versus large communities.

2.2 The model

Consider a population of N individuals (we will use the term "community") with opinions on a set of J topics or issues. Each individual is represented as a J-dimensional binary vector with components in $\{0, 1\}$ which we will call an *opinion type*. There is a J-dimensional vector, $w = (\omega_1, \ldots, \omega_J)$, whose components are the positive weights placed on each issue. This weight vector represents the way in which the community weighs the importance of the issues against each other. There is also a threshold parameter $\alpha \in [0, 1]$. The agreement level $A_{x,y}$ between any two individuals x, y is the weighted average of their agreements on all the issues. Let us denote the binary opinion vectors of the two individuals as $(\sigma_1^{(x)}, \ldots, \sigma_J^{(x)})$ and $(\sigma_1^{(y)}, \ldots, \sigma_J^{(y)})$. We define

$$A_{x,y} = \frac{\sum_{j=1}^{J} \omega_j \delta_{\sigma_j^{(x)}, \sigma_j^{(y)}}}{\sum_{j=1}^{J} \omega_j}, \tag{2.1}$$

where $\delta_{a,b}$ is the Kronecker symbol: $\delta_{a,b} = 1$ if $a = b$ and $\delta_{a,b} = 0$ otherwise. At each discrete time step a random individual is chosen to be active, and an "interlocutor" is randomly picked among the rest of the community, as well as a random issue j. If the agreement level is sufficiently high, such that $A_{x,y} \geq \alpha$, then the active individual changes their opinion on issue j to match the interlocutor, if it is currently different. We will call this a *type one* change. If on the other hand, the agreement is sufficiently low, $A_{x,y} < \alpha$, then the active individual changes their opinion on issue j to be different from the interlocutor, if it is currently the same. We will call this a *type two* change. In all other cases no changes occur. These dynamics capture the idea that relative similarity leads to greater similarity, and relative dissimilarity leads to greater dissimilarity. The threshold parameter α defines what level of agreement is considered "low" or "high".

The absorbing states of this system have the following structure. The population consists of two subgroups of individuals of sizes N_1 and N_2 respectively, with $0 \leq N_1 \leq N$ and

$N_2 = N - N_1$. In both subgroups (which we call "factions"), all individuals agree with each other on all of the issues. The opinion vectors of the two factions are complimentary (that is, they differ on all of the issues). If $0 < N_1 < N$, we call the equilibrium a two-faction equilibrium. Otherwise, we call it a one-faction equilibrium. All absorbing states (or "final states") are characterized by an $N \times N$ agreement matrix (whose elements are pairwise agreements), which consists entirely of $0's$ and $1's$.

It is clear that once in a one-faction equilibrium, no further change is possible under the rules described above, because $A_{x,y} = 1$ for all pairs of individuals, and the only action would be to change opinions to become more similar, which is impossible because they are already identical. Similarly, in a two-faction equilibrium, no further change is possible. This is because if two interacting individuals belong to the same faction, no action is taken as explained above; if two interacting individuals belong to different factions, their agreement is $A_{x,y} = 0$, and the only possible action of the active individual is to switch an issue to differ from the interlocuter, but all the issues are already different, so no action is taken. Finally, we can show that no other fixed point is possible. Let us suppose for example that there are three individuals in the population such that their opinion types are different. Then there will be a pair of individuals whose agreement score is $0 < A_{x,y} < 1$, such that they agree on some issues and disagree on others. In this case, if $A_{x,y} \geq \alpha$ and j is an issue where the pair disagrees, a change towards agreement may take place; if $A_{x,y} < \alpha$ and j is an issue where the pair agrees, a change towards disagreement may take place. This shows that there are nonzero transition probabilities away from this state, and therefore it is not absorbing.

The simulations proceed as follows. The weight ω_J on the Jth issue will always be set equal to 1 and serves as a reference weight for the other issues. A parameter set $(\omega_1, \ldots, \omega_{J-1}, \alpha)$ is chosen, which consists of the weights on the other $J-1$ issues, and the value of α. Each individual begins as a randomly generated J-dimensional vector with components taking

values in $\{0, 1\}$. After a meeting has occurred and the vectors are updated, the system checks to see if the final state has been reached (by checking that the agreement matrix consists of only zeros and ones). If a final state has been reached, the simulation ends; otherwise the simulation continues. The process of generating meetings until the system reaches a final state will be called a *trial*.

2.3 Dynamics of faction formation with weighted issues

2.3.1 One weighted issue: patterns of behavior

Let us assume that the weighting vector takes the form

$$(\omega_1, \omega_2, \ldots, \omega_J) = (X, 1, \ldots, 1),$$

where X is a positive real number. The effect of $0 < X < 1$ is to model a situation where the weighted issue is not as important as the others. $X > 1$ models a situation where one issue is more important than the others, and the case $X = 1$ corresponds to a situation where all issues are equally weighted, similar to what has been done before in variations of the Axelrod model.

We first take the number of issues to be $J = 4$, such that the weights are given by $(X, 1, 1, 1)$. We are interested in the long-term behavior of the system, so for each value of X, we run multiple trials and record the statistics of the following measures:

1. The size of the larger faction at the final state,

2. The number of factions at the final state (which is related to (1)), and

3. The number of meetings required to reach the final state.

If we fix the threshold parameter α, we observe that the mean largest faction size depends on X in a stepwise manner, and for some values of α are non-monotonic in X, see figure 2.1(a). The other characteristics such as the convergence time and the proportion of runs that result in two factions, are also stepwise-constant functions of X (not shown).

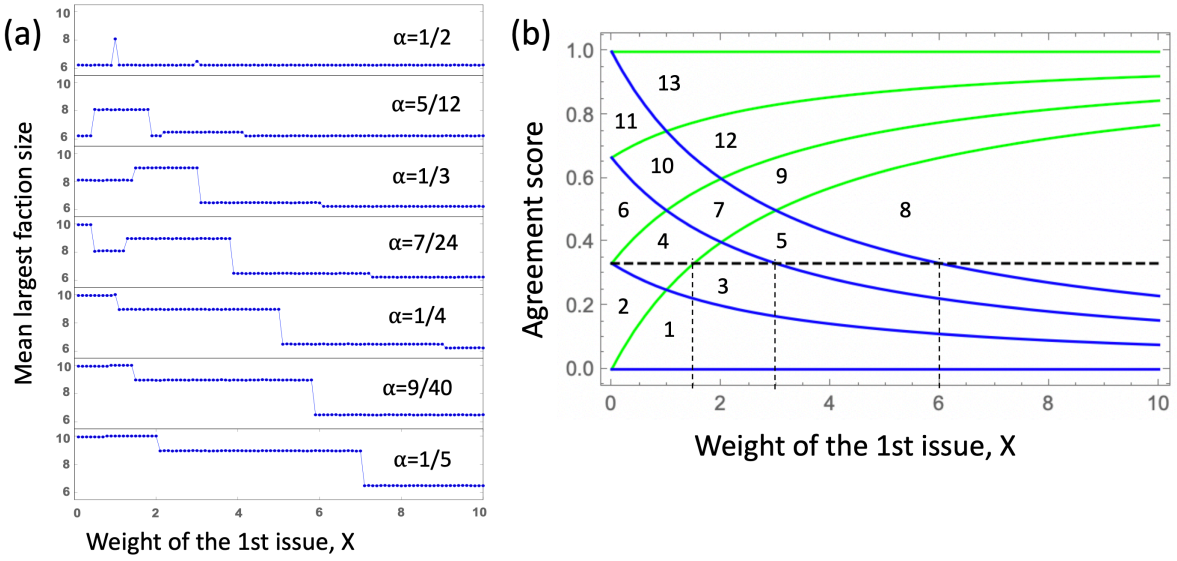


Figure 2.1: (a) The mean largest faction size evaluated over 20,000 trials, plotted as a function of X (the weight of the first issue), for several values of α, with $N = 10, J = 4$. (b) Agreement scores for different configurations of individuals, with $J = 4$, plotted as a function of X. Blue: functions $a_{00}, a_{01}, a_{02}, a_{03}$ (equations (2.2)); green: functions $a_{10}, a_{11}, a_{12}, a_{13}$ (equations (2.3)). The horizontal dashed line denotes the level $1/3$; the vertical dashed lines show at which values of X the line $\alpha = 1/3$ intersects the a_{ij} lines. The numbers between the agreement score curves are the region labels used in simulations.

This behavior can be explained by noting that the agreement functions for a fixed value of X, can take only a small number of values, depending on the set of issues with coinciding (and opposing) opinions. Let us denote by $a_{ij}(X)$ the agreement function corresponding to the situation where i weighted opinions and j unweighted opinions coincide (in our case, $0 \leq i \leq 1, 0 \leq j \leq 3$). If all the opinions between two individuals are different, the agreement function is $a_{00}(X) = 0$. If 1, 2, or 3 opinions coincide, all of which are

non-weighted (not opinion 1), then the agreement functions are given by

$$a_{01}(X) = \frac{1}{X+3}, \quad a_{02}(X) = \frac{2}{X+3}, \quad a_{03}(X) = \frac{3}{X+3}, \tag{2.2}$$

respectively. If 1, 2, or 3 opinions coincide such that one of them is weighted (opinion 1), then the agreement functions are given by

$$a_{10}(X) = \frac{X}{X+3}, \quad a_{11}(X) = \frac{X+1}{X+3}, \quad a_{12}(X) = \frac{X+2}{X+3}. \tag{2.3}$$

Finally if all of the opinions between two individuals are the same, then the agreement function is $a_{13}(X) = 1$. These 8 functions partition the space (X, α) into 13 regions (see figure 2.1(b)). Observe that given any point (X, α), for each pair (i, j), either $a_{ij} \geq \alpha$ or $a_{ij} < \alpha,\ i, j = 0, \ldots, 3$. Since the agreement functions determine the switching behavior of the model given fixed (X, α), we conclude that for any points (X, α) in the same region the long-term behavior of the model will be (statistically) the same. To illustrate this, take the example of $\alpha = 1/3$ (the dashed horizontal line in figure 2.1(b)). Observe that when $0 < X < 3/2$ then the model is in region 4; then for $X \in [3/2, 3]$ the model is in region 3; for $X \in (3, 6]$ it is in region 5, and finally for $X > 6$ it is in region 8. Looking at the simulation results in figure 2.1(a) corresponding to $\alpha = 1/3$, we observe that the behavior changes in a step-wise manner as X changes through these region boundaries; the behavior of the mean largest faction size remains constant inside each of these regions.

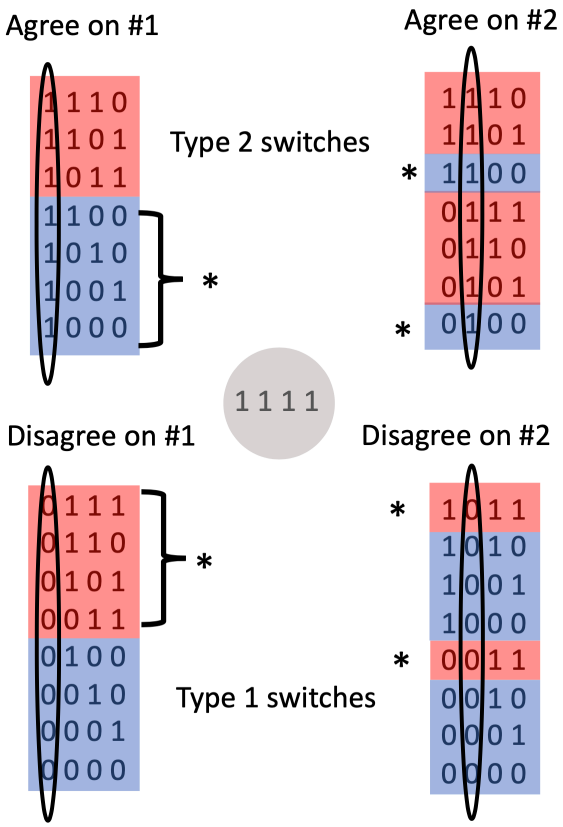


Figure 2.2: An example demonstrating the counting algorithm to determine how many different interlocutor types can lead to a type 1 and type 2 changes in an active individual of type (1111) (center). Here we assume that the system is in zone 6. Types with agreement score that is "high" are shaded red and those with a "low" agreement score are shaded blue. See text for details.

2.3.2 Agreement-disagreement score

In figure 2.3 we present the mean time to fixation, the two faction proportion, and the mean largest faction size as heat plots in the $X - \alpha$ space. If $X = 1$, all three quantities are monotonic functions of α: time to fixation and the proportion of two-faction solutions both increase with α, and the mean largest faction size decreases with α. When weighting is added to one of the issues, patterns become more complex. While all three characteristics are clearly (in a statistical sense) determined by the region, the dependence on X and α is non-monotonic and difficult to interpret. In order to make sense of the inter-regional variability of behavior, we have implemented the following procedure of quantifying the

switching behavior of the model.

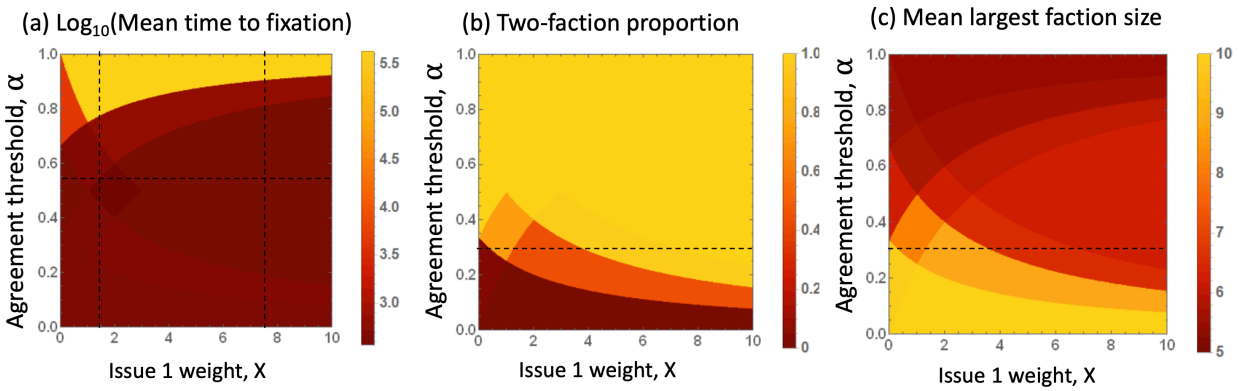


Figure 2.3: A model where $J = 4$ with a single weighted issue: the mean time to fixation (a), two faction proportion (b), and mean largest faction size (c) are shown as heat plots in the $X - \alpha$ space. The dashed lines help notice the non-monotonic behavior with respect to X and α.

We are interested in the long-term behavior of the model, randomized over different initial conditions. Observe that in the current setting ($J = 4$ issues) there are $2^4 = 16$ different opinion types. Without loss of generality, let us suppose that an active individual has opinion string $(1, 1, 1, 1)$ (figure 2.2, the middle circle), and count how many interlocutor configurations can result in different types of switches. We are interested in type 1 switches (that is, switches to agree), and type 2 switches (switches to disagree). In figure 2.2 we illustrate how to quantify the switching behavior by using a specific example of region 6. This region (see figure 2.1(b)) is defined by the following inequalities:

$$a_{01}(X) < \alpha \leq a_{02}(X) \quad a_{11}(X) < \alpha \leq a_{12}(X).$$

These inequalities state that in order to score a “high” level of agreement, an interlocutor must coincide with the active individual on at least two non-weighted issues.

Let us first count how many types of interlocutor will cause the active individual to perform a type 1 change on the weighted issue. This will happen if (1) the interlocutor disagrees with the active individual on the weighted issue, and (2) the agreement score qualifies as

"high". In figure 2.2, all types that disagree with the active individual on the 1st issue are listed in the bottom left region, and the ones with a high agreement score (for region 6) are colored red. There are four such types, meaning that there are four type 1 changes for issue 1.

To count type 2 changes for issue 1, we note that this type of switch happens if (1) the interlocutor agrees with the active individual on the weighted issue (left top region in figure 2.2), and (2) the agreement score qualifies as "low" (blue shade). Again, there are four such types, meaning that there are four type 2 changes for issue 1.

Similarly, to count type 1 switches on issue 2 (or any non-weighted issues), we select types that disagree with the active individual on issue 2 (bottom right) and have a high agreement score (red), resulting in 2 possibilities. For type 2 switches on issue 2, we select types that agree with the active individual on issue 2 (top right) and have a low agreement score (blue), resulting in 2 possibilities. Since the same calculation applies to any of the 3 non-weighted issues, we conclude that there are 2×3 type 1 changes and 2×3 type 2 changes on non-weighted issues. Adding everything up, we can see that there are a total of 10 type 1 and 10 type 2 changes for region 6.

The number of type 1 and type 2 changes depends only on what region (figure 2.1(b)) the system is in. A composite measure that turns out to be a very useful correlate for the system's long term dynamics is what we call "Agreement-Disagreement Score" (ADS),

$$\delta = \frac{\text{total type 1 changes} - \text{total type 2 changes}}{\text{total type 1 changes} + \text{total type 2 changes}} \tag{2.4}$$

Table 2.1 presents the numbers of type 1 and type 2 changes as well as ADS for all the regions in figure 2.1(b) (see also table A.1 in the appendix for more details).

By definition, ADS satisfies $\delta \in [-1, 1]$. It quantifies the behavior of the model by giving

Region	Total type 1 switches	Total type 2 switches	ADS, δ
1	28	0	1
2	25	1	0.92
3	19	3	0.73
4	16	4	0.60
5	13	9	0.18
6	10	10	0
7	10	10	0
8	12	12	0
9	9	13	-0.18
10	4	16	-0.60
11	1	25	-0.92
12	3	19	-0.73
13	0	28	-1

Table 2.1: The switch configurations for different regions, see figure 2.1(b).

the relative proportion of type 2 changes to type 1 changes. Values of δ close to 1 indicate that in such a region the proportion of type 2 changes to type 1 changes is small, that is, the population can be characterized as being "trusting," and individuals are often willing to change their opinions to agree with their interlocutors. When δ is near -1, the switch behavior consists of mostly type 2 switches, which means that more often than not, individuals view others as "opponents" and are willing to switch their own behavior just to be different.

In figure 2.4 we show how the three system features of interest (the mean time to fixation, the two-faction proportion, and the mean largest faction size) depend on the ADS, δ. The results for a system with a single weighted issue are depicted by blue points. Let us first examine the probability to end up with two factions (figure 2.4(a)) and the mean larger faction size (figure 2.4(b)). For negative values of δ (which corresponds to having more ways to disagree than to agree with an interlocutor), all simulations always result in two factions, and the system is more balanced (similar faction size) for larger negative ADS. Increasing values of δ from 0 to 1 leads to a decay of the probability of the two-faction outcome and the existence of one dominant faction. Region 1, which has $\delta = 1$ is char-

acterized by a 100% convergence to one faction, because the only way to switch in this system is to agree with the interlocutor.

Next we turn to the graph of the mean convergence time, figure 2.4(c). We observe that time to fixation increases dramatically as $\delta \to -1$ (please note the log scale on the figure). This is because for negative values of δ, individuals tend to switch away from each other, which in contrast to switching towards each other, precludes agreement, and thus delays convergence. However, this does not imply in general that greater proportions of type 1 changes correspond to shorter convergence times. Note that the graph in figure 2.4(c) has a shallow minimum at $\delta = 0$, such that convergence takes a slightly shorter time for systems that are balanced (have the same number of type 1 and type 2 switches) compared to say region 1, where individuals can only switch to agree. This can be explained by noticing that having $\delta = 1$ "forces" the system to converge to a single-faction equilibrium, and the dynamics can only stop when a one-faction system is achieved. Since on average, such a state is farther from the (random) initial condition than a two-party system, this trades-off to slightly increase the convergence time.

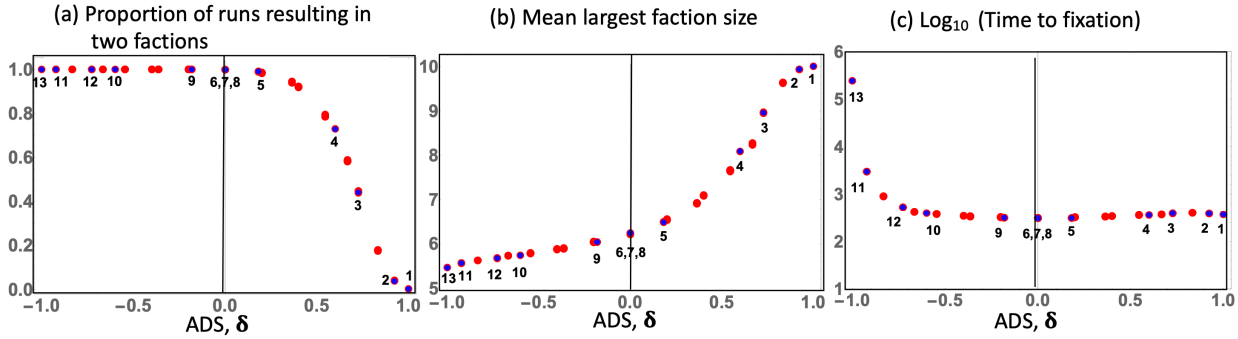


Figure 2.4: The two faction proportion (a), mean largest faction size (b), and the mean time to fixation (c), plotted against δ. Blue points correspond to a system with a single weighted issue; each blue point is labeled with the region is represents. Red points correspond to the more general system where all issues could have a different weight. Other parameters are $N = 10, J = 4$. Each point represents the mean over 20,000 independent simulations. Standard errors are too small to be visible; see figure A.2 for standard deviation bars. See figure A.1 that shows similar results for a system of a larger size.

2.3.3 Multiple weighted issues

Here we will generalize our findings by allowing more than one of the weights to vary. Let s be a binary string of length J: $(s_1, \ldots, s_J)$. For each string s we define

$$a_s(\omega_1, \ldots, \omega_{J-1}) = \frac{\sum_{j=1}^{J-1} s_j \omega_j + s_J}{\sum_{j=1}^{J-1} \omega_j + 1},$$

a function of $J-1$ variables $\omega_1, \ldots, \omega_{J-1}$. For fixed values of $\omega_1, \ldots, \omega_{J-1}$, these functions for all binary vectors s represent all possible agreement scores that can exist in the population. The 2^J functions $a_s(\omega_1, \ldots, \omega_{J-1})$ partition $(0, \infty)^{J-1} \times [0, 1]$, and each parameter set $(\omega_1, \ldots, \omega_{J-1}, \alpha)$ will lie inside a region bounded by some of the functions a_s. The statistical properties of the system with fixed parameters are uniquely defined by which region the parameter set $(\omega_1, \ldots, \omega_{J-1}, \alpha)$ belongs to.

As an example, consider the case with $J = 4$ issues, where all issues might have a different weight. We fix the weight on the 4th issue at $\omega_4 = 1$ to serve as a reference weight and allow $\omega_j,\ j = 1, 2, 3$ to vary. Now the agreement functions

$$a_s(\omega_1, \omega_2, \omega_3) = \frac{\sum_{j=1}^{3} s_j \omega_j + s_4}{\sum_{j=1}^{3} \omega_j + 1}$$

partition $(0, \infty)^3 \times [0, 1]$ into distinct regions. In each of these regions, as we observed above, the numbers of type 1 and type 2 changes is constant from point to point, and δ is constant inside each region. In the system with 4 issues each of which could have a different weight, we have a total of 148 different regions, which are difficult to visualize.

Instead, we can calculate the ADS for each of the regions, which creates a one-dimensional parameter to characterize the system's properties (see Appendix A for details). Figure 2.4 (red dots) plots the two-faction proportion, the mean largest faction size, and the mean time to fixation, as a function of ADS for this more general system. We observe that the

tendencies are exactly the same as those identified in a previous system (which comprises a subset of different cases in the more general system). In particular, relatively more type 2 opinion switching corresponds to longer convergence times and a greater proportion of fixed states resulting in 2 factions, while the mean largest faction size in the fixed state tends to be smaller.

2.3.4 Individual weight distributions

Finally, we modify the above framework to include individual weight distributions. For several different fixed values of α, we assumed that individuals in the population each have their own weighting scheme; those were picked at random and independently from the uniform distribution (5 different choices of random weighting schemes per α). This represents the model where relative importance of issues could vary from person to person. Results of such simulations are presented in figure 2.5 (see also figure A.2 for larger population size simulations).

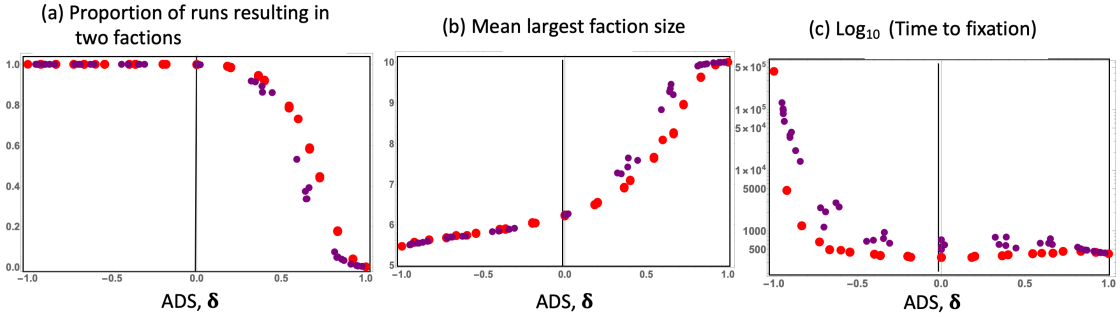


Figure 2.5: Individual weighting of the issues. Same as figure 2.4, but showing results for simulations where each person assigned weights individually (purple). The horizontal axis is the population average of individual ADS values. Plotted are: the two faction proportion (a), mean largest faction size (b), and the mean time to fixation (c), plotted against δ. As in figure 2.4, red points correspond to weights being the same for all individuals in the population (all weighted issues). Other parameters are $N = 10$; each point represents the mean over 20,000 independent simulations. Standard error is too small to be visible on the graph.

In this model, we calculated a "personal" ADS, δ_i, for each individual i for $1 \leq i \leq N$, by

using the algorithm described in figure 2.2, and the found the population ADS by averaging these values over all the individuals. Figure 2.5 plots the population dynamic observables (proportion of runs resulting in two factions, mean largest faction size, and time to fixation) as functions of the ADS averaged over the population. We can see that the results are very similar (although not identical) to the case where uniform weighting was implemented for all individuals in the population. We observe that despite heterogeneity in issue weights, the mean ADS places the system on the spectrum from most "cautious" to most "trusting", which can approximately predict the characteristics of the population dynamics.

Apart from the mean ADS, it is instructive to consider individual ADS values and determine if they influence the expected behavior of the individuals in the population. To investigate this, we traced each individual to see if they ended up in the larger of the two factions (including the situation when only a single faction was formed or the two factions were exactly the same size; both of these ambiguous outcomes were counted as "belonging to the larger of the factions"). Figure 2.6(a) plots the probability for each person to end up in the largest faction, as a function of the individual value of δ. These are grouped and color-coded by the α-value. We find a statistically significant trend that the probability of joining the larger of the groups grows with the individual ADS. In other words, individuals that are more trusting are more likely to join the majority group compared to those that are more cautious.

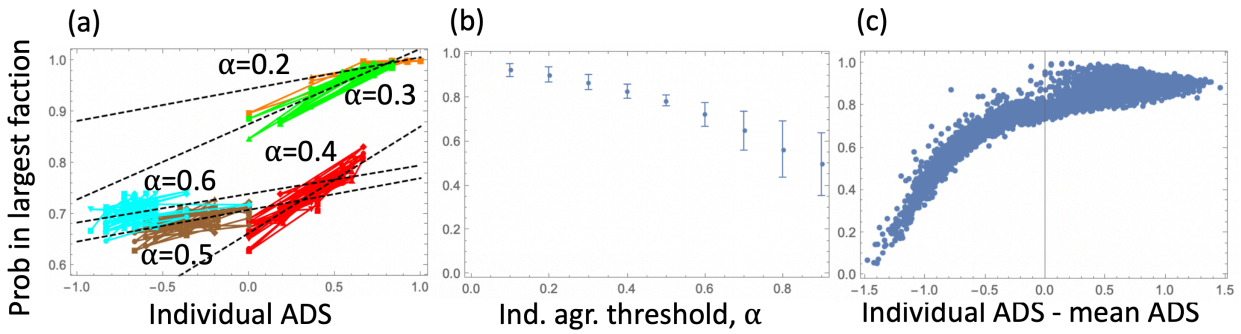


Figure 2.6: Opinion dynamics in heterogeneous populations. (a) Individual weights, common agreement threshold: The probability for individuals to join the majority group is plotted against the individual ADS values, grouped and color-coded by the value of α. The linear fit for each set is presented by dashed lines; in all these cases $p < 10^{-3}$; $N = 20$, 1000 simulations per parameter set. (b) Individual agreement thresholds and weights: The probability for individuals to join the majority group is plotted against the individual agreement threshold values, α (the average and the standard deviations are presented for each α). (c) The same as (b) but plotted as a function of the difference between the individual and the population ADS. The rest of the parameters are the same as in figure A.2. For (b),and (c), $N = 10$, 400 parameter combinations were used, 1000 simulations each.

Finally, we considered the highest degree of population heterogeneity where not only the weighting of the issues, but the agreement thresholds were different for different individuals. In these simulations, the individual α values were chosen uniformly from a discrete set of values $\{0.1, 0.2, \ldots, 0.9\}$, and the issue weights, as before, were chosen from a uniform probability distribution in $[0, 1]$. Figure 2.6(b) groups the simulations results by the α values and shows the mean probability of individuals to end up in the larger of the factions (the error bars show the standard deviation). We observe that larger individual agreement threshold values result in a lower probability to join the majority. A similar result is shown in panel (c) of figure 2.6, where the same data are replotted differently. Instead of grouping individuals by their agreement thresholds, we plotted the probability to end up in the larger faction as a function of the δ-difference, which is the individual ADS of each individual minus the population average of the ADS. In this graph, the horizontal axis informs us on how “trusting” or “cautious” each individual was compared to the other members of

the same population. We can see that individuals that are more trusting than their peers (positive ADS difference) are more likely to join the majority compared to those whose ADS is lower than the average.

2.4 Opinion dynamics in the presence of an influencer (small vs. large communities)

Next, we study what happens when the community is affected by the presence of an outsider who we will call the *influencer*. We will also see how a larger community ($N = 50,\ N = 100$) effects the results, both with, and without the presence of an influencer. As with the other members of the community, the influencer has a binary opinion on each of the J issues. In contrast to regular community members, the influencer never changes their opinions (that is, they are never chosen as an active individual), but instead they play the role of an interlocutor disproportionately often. At each meeting time, in the absence of an influencer, the chance that each regular individual is chosen as the interlocutor to an active individual is given by $1/(N-1)$. We assume that the influencer has a probability p to be chosen as the interlocutor. We will call this the *influence probability* The probability of the rest of the individuals to be chosen is uniform and scales accordingly: for influence probability p the probability of any other community member to be chosen is $(1-p)/(N-1)$. The case of $p = 0$ corresponds to the dynamics in the absence of an influencer.

Since the influencer's opinions never change, while other community members' opinions can switch, the presence of an influencer significantly reduces the space of possibilities for the final state: the system must converge to a fixed state, in which one faction agrees with the influencer on all the issues, and one faction disagrees with the influencer on all the issues. One of these factions may have size zero.

2.4.1 Faction sizes

First, we study which parameter values result in the largest faction having the same opinion vector as the influencer. In figure 2.7(a) we depict the proportion of trials in which the largest faction matches the influencer against the variable δ for influence probabilities $0.1,\ 0.4,\ 0.7,\ 1.0$. We observe that at each influencer probability the plots monotonically increase with δ. When the community is more likely to switch to agree (disagree), then anytime the active individual meets with the influencer, they are more likely to switch to agree (disagree) with the influencer. Larger influence probabilities amplify this effect. At $p = 0.1$, the proportion of runs where the larger faction matched the source is about 0.5 at $\delta = -1$ and increases to 1 at $\delta = 1$. At $p = 0.4$ this proportion begins at 0 and increases in an approximately linear fashion to 1. At levels greater than $p = 0.4$ the match proportion has a more "logistic" appearance. For $p = 1.0$, active individuals can only have the influencer as their interlocutor, which corresponds to the extreme case where the whole population converses with a single source of influence. For δ near -1, the presence of this extremely influential opinion source, however, has a negative effect, because the population always converges to a set of opinions that is a polar opposite of those of the influencer.

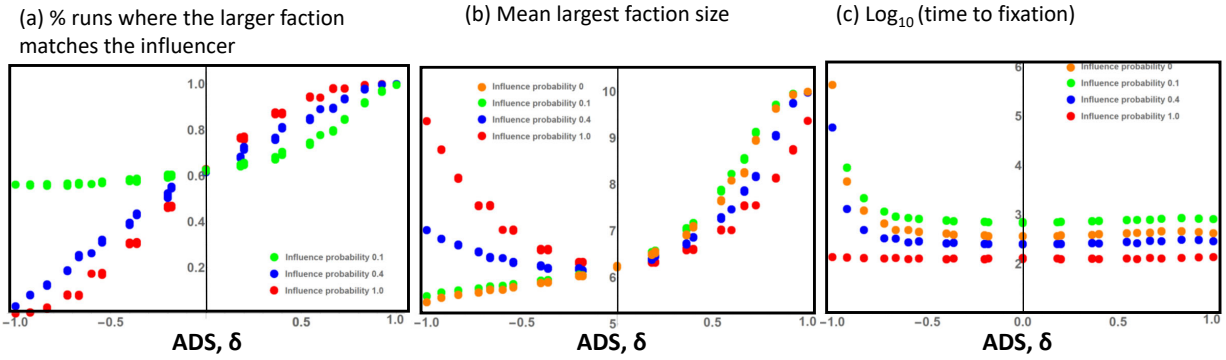


Figure 2.7: Opinion dynamics in the presence of an influencer. (a) The fraction of runs where the largest faction matches the opinion vector of the influencer; (b) the mean largest faction size; (c) $\log_{10}$ of the time to fixation. All of these measures are presented as functions of ADS, δ, and the different colors represent the influencer level. Parameters are $N = 10, J = 4$, all issues could be weighed.

Consistent with these observations, the dependence of the mean largest faction size on δ changes as we change the influence probability p. Figure 2.7(b) shows that for small values of p, the mean largest faction size is a monotonically increasing function of δ, much like was the case in the absence of an influencer. As p increases, however, the mean largest faction size becomes larger for negative and smaller for positive values of δ. For the highest influence probability, it is a symmetric function of δ, with the very unbalanced, or a 1-faction, steady state for $\delta = \pm 1$. This means that a strong influencer significantly changes the opinion dynamics of a community with a strong tendency to disagree (large negative δ): in the absence of an influencer, such a community tends to converge to a balanced two-faction system, but the presence of a strong influencer leads to the creation of a single faction with opinions that are opposite of those of the influencer. In other words, in the case of large negative ADS, an influencer tends to unite the community against itself.

2.4.2 Time to fixation

In this section, we study how the time to fixation changes as the influence probability increases, see figure 2.7(c). We observe that as we first introduce a low-p influence, the convergence time at all values of δ increases compared to the system in the absence of an influencer. As the influence level becomes higher, however, the convergence time drops significantly and becomes shorter compared to that in the absence of an influencer.

These trends have a clear intuitive explanation. First let us compare a population without an influencer with a population with a influence probability $p = 0.1$. In the latter case, the probability to meet with an influencer is as high as to meet with any other person. Without an influencer, the population will converge to a steady state which, statistically speaking, is close to its initial opinion distribution. In the presence of an influencer, however, the system will not stop changing until the two factions are the one coinciding with the influencer and the opposite of this, which for most initial conditions is a state that is fartherfrom the initial

opinion distribution, and thus takes longer to reach.

If, on the other hand, the influence probability is very high, this leads to a different tendency: because of frequent meetings with the influencer, a large proportion of the changes will point in the same direction, and the population will relatively quickly reach convergence to a state defined by the influencer (or is opposite to that of the influencer). This explains the drop in the convergence time as p increases. It is tempting to think that time to fixation maximizes at p satisfying

$$\frac{1-p}{N-1} = p. \tag{2.5}$$

So in this case where $N = 10$ we would expect the time to fixation at $p = 0.1$ to be the maximum. However, we must recall that there may be other individuals with the same opinion type as the influencer and a meeting with any one of these members of the community. Therefore, the value of p satisfying (2.5) is likely only a good guideline for finding a maximum time to fixation due to the randomness of the system and the value of δ since, as we have seen, the number of persons of the same type as the influencer depends on δ. We will look into this more closely in the next section.

2.4.3 Larger communities

Lastly, we study how communities of larger populations affect the results we have seen above, with and without an influencer. We will look at communities of sizes $N = 100$ and $N = 50$. Due to extremely long convergence times for some values of δ at these population sizes, we adopt the following conventions for this section: (1) the simulations have an additional stop condition: if 10^7 meetings occur without reaching a fixed state, the simulation stops; and (2) we record the number of meetings at the end of a simulation whether or not convergence occurs, but we do not record the mean largest faction size,

nor whether the larger faction matches the influencer.

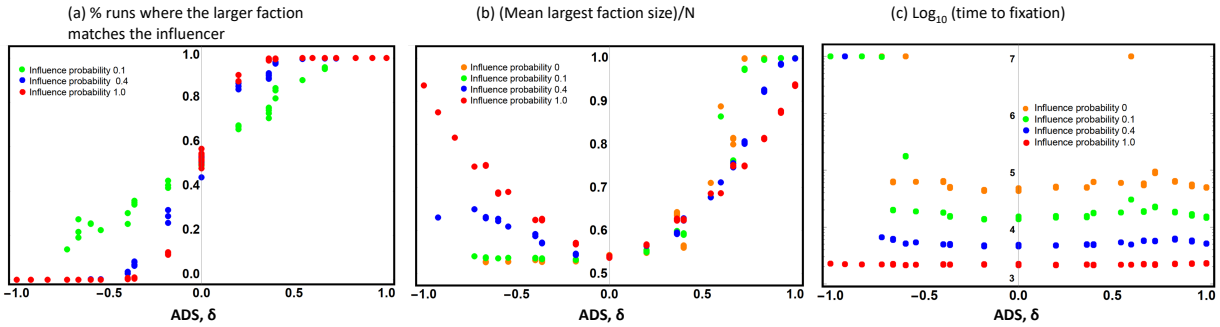


Figure 2.8: Opinion dynamics in the presence of an influencer. (a) The fraction of runs where the largest faction matches the opinion vector of the influencer; (b) The mean largest faction size/N; (c) $\log_{10}$ of the time to fixation. All of these measures are presented as functions of ADS, δ, and the different colors represent the influence probability. Parameters are $N = 100, J = 4$, all issues could be weighed, and values are based on 500 trials.

We see that figures 2.8(a) and (b) have the same general shape as in those of figures 2.7(a) and (b). The differences can be accounted for by the much larger variability in considering 500 trials per point as compared to 20,000 trials, as well as the conventions adopted for this section mentioned above.

The difference in time to fixation, figure 2.8(c) is more curious. Observe that at $\delta = 0.6,\ -0.6$ there are jumps when the influence level is low. We will analyze the reasons for this in chapter 2. However, larger influence probabilities rapidly diminish this effect. We also observe that for the given influence probabilities $(0,\ 0.1,\ 0.4,\ 1.0)$ we do not see time to fixation appearing to increase before monotonically decreasing in the left pane. However, this is due to $p = 0.1$ being larger than $1/100$ by a factor of 10. In figure 2.9 we see in the left pane that time to fixation appears to have a maximum in the influencer probability interval of $[0.005, 0.015]$ and monotonically decreases thereafter. We also observe similar behavior when $\delta = 0.60$ in the right pane of figure 2.9. Here, we use $N = 50$ so that the system reaches a fixed state prior to 10^7 meetings. The maximum occurs in $[0.015, 0.025]$. Time to fixation tends to decrease monotonically from there. This can also be seen in figure 2.8(c) by comparing those values of δ for which the time to fixation reaches the

imposed threshold of 10^7 against the others.

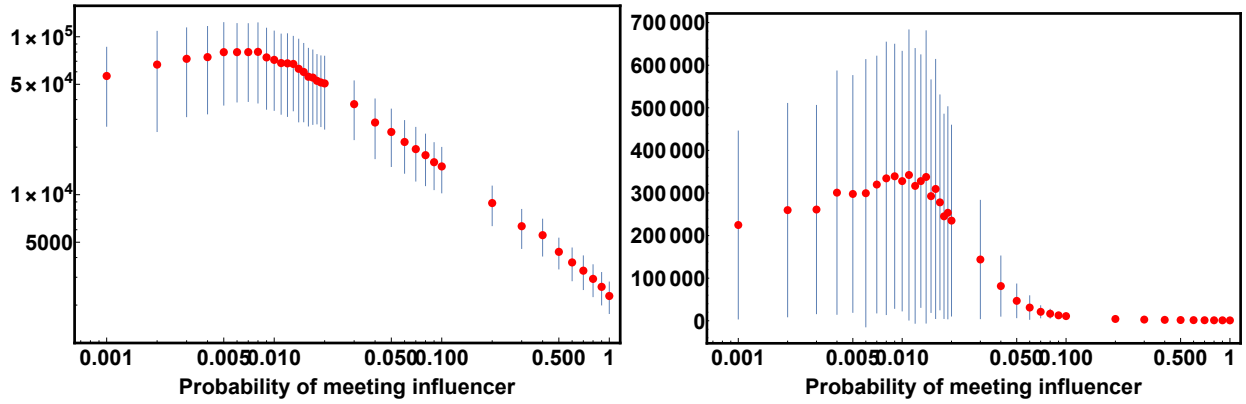


Figure 2.9: Mean time to fixation in the presence of an influencer.
Left pane: $N = 100$, $\delta = 1$, Log-Log scale. Right pane: $N = 50$, $\delta = 0.6$, Log-Linear scale. Point values are based on 500 trials per point. Error bars depict standard deviation. In some cases standard deviation is too small for the error bar to appear.

2.5 Discussion

We considered a stylized model of opinion dynamics, where people are characterized by a string of J binary opinions. They interact in random pairs, and change their binary opinions according to the agreement score, A, that they have with their interlocutor. If the agreement score is relatively high (above a certain threshold, $A \geq \alpha$), they swap a randomly chosen opinion to coincide with their interlocutor. If the agreement score is lower than the threshold, $A < \alpha$, they swap a randomly chosen opinion away from their interlocutor.

In this system, only two long-term outcomes are possible: either all the individuals become identical in all the opinions, or they form exactly 2 groups of shared opinions, with all the opinions exactly opposite between the groups. If all the issues are weighted in the same way (that is, they are equally important for the calculation of the agreement score), an increase in agreement threshold (parameter α) increases the likelihood of the two-faction outcome. If, however, one of the issues is more important than the rest, more complex pat-

terns are observed, and the probability of the two-faction outcome, the mean largest party size and the time to fixation become non-monotonic functions of the agreement threshold, α, and/or issue weights.

In order to explain the variety of observations, it is helpful to define a scalar parameter, δ, which we called the agreement-disagreement score (ADS). This score depends on both the issue weighting and the agreement threshold. It is calculated by counting the number of different types of interlocutors that will cause a target individual to switch toward the interlocutor (type 1 switch) and to switch away from the interlocutor (type 2 switch). Then the proportions of type 1 and type 2 changes is calculated and the difference between the two gives the agreement-disagreement score. This is a quantity that can range between -1 and 1, with values near 1 corresponding to situations where individuals tend to switch to agree most of the type, signaling a relatively low agreement threshold and a "friendly" or "trusting" disposition of the community. Values of δ near -1 correspond to the opposite situation where individuals tend to switch away, to increase the difference with the interlocutors, which represents a very "unfriendly" or "cautious" community.

In the absence of issue weighting (that is, all weights are equal), agreement-disagreement score is a monotonically decaying function of the threshold parameter α (and thus is redundant). In the presence of issue weighting however it provides a scalar parameter that places a system on an interval $[-1, 1]$ and allows for predicting important observables of the behavior, such as convergence time and faction structure. For example, for societies with negative δ, a two-faction outcome is a certainty. For positive δ, the probability of a one-party outcome gradually increases and reaches 1 with $\delta = 1$. The time to fixation increases exponentially as δ approaches -1.

In communities with external influence, the fixed state is determined by the influencer's opinions - any faction in the steady state will agree with the influencer on all the issues, or disagree on all the issues. In a community with negative δ values (that is, a "cautious" com-

munity), the largest faction tends to consist of individuals that disagree with the influencer, and the tendency is the opposite for positive-δ, "trusting" communities. The size of the effect clearly grows with the influencer level (that is, with the frequency of communication with the influencer). Further, the population faction structure is affected by the presence and the level of the influencer. In the absence of an influencer, negative δ values are associated with two balanced factions in steady state, and an increase in $\delta > 0$ leads to an increase on one-faction outcomes (and a larger mean largest faction size). Increasing the level of an influencer breaks this pattern and leads to an increasing frequency of one-party outcomes for negative δ values, especially for δ close to -1, such that a more symmetric picture emerges: both for very "trusting" ($\delta \approx 1$) and very "cautious" ($\delta \approx -1$) populations, a single party tends to emerge, but in the former case it coincides with the influencer's opinions, while in the latter case it is characterized by the opposite set of opinions. Intermediate populations (δ near zero) retain the same mean largest faction size regardless of the influence level.

Time to fixation again is significantly affected by the presence of an influencer, especially for negative-δ population and also at values of δ at which larger populations have extremely long convergence times, until the influencer power is large enough that time to fixation no longer experiences such a large increase as $\delta \to -1$. One can say that the presence of an influencer unifies populations to align with the source of influence if $\delta > 1$ and to disagree with it if $\delta < 1$, and consensus is reached relatively fast for both extremely "trusting" and extremely "cautious" populations. Furthermore, without a community influencer, larger populations greatly affected the shape of the time to fixation plot at certain points. As mentioned above, we will study this phenomenon in Chapter 2.

Finally, we investigated further inhomogeneities of populations, such as different weights assigned to issues by different individuals. It was found that the population mean ADS continues to be a reasonably good proxy for the opinion dynamics, where large negative

ADS values predict very long fixation times and an almost certain balanced two-faction outcome, whereas large positive ADS values result in a quicker convergence and a single-faction steady state. Moreover, individual ADS values reflect individual behavior. Agents with relatively higher δ-values are more likely to end up in the majority faction whereas more "cautious" (lower δ) individuals are less likely to join the larger faction.

This study is a first step toward understanding complexities of social opinion dynamics associated with variable importance, or "weight", of different issues that comprise individuals' opinion profile. Even though populations characterized by different agreement thresholds (α) and different weights associated with the J issues "live" in a J-dimensional parameter space ($J-1$ dimensions for the issue weights that sum up to one, and an extra dimension for α), it was possible to find a simple index-type variable that described the behavior of such populations. This variable had the power to predict the steady state solution statistics and structure and convergence type. Extensions of this work will include interaction networks other than a complete graph, e.g. spatial networks, or networks where individuals are more likely to interact with those who are similar to them (homophily). Further, it would be interesting to investigate the effect of multiple influencers.

Chapter 3

The ODE theory of weighted opinion dynamics

3.1 Introduction

In chapter 2 we introduced and studied a weighted opinion dynamics model which used both agreement and disagreement dynamics. In the current chapter, we study a deterministic system of ODE's whose discrete/stochastic analog is said model. We demonstrate that the deterministic system has a neutrally stable solution which exists for all values of the parameters N, J, α, and weights. We argue that if this solution is the only stable one, the behavior in the corresponding stochastic system leads to extremely long times to a fixed state. Furthermore, we demonstrate that the cases observed in the previous chapter (δ near -1, and for $J = 4$, $\delta = \pm 0.6$ with larger populations) are precisely those in which this solution exists as the only stable solution. We also give an algorithm for finding solutions to this system when α is small relative to the possible agreement scores implied by the issue weights.

3.2 The set-up

Suppose in a community of N individuals we have J binary opinions so that the total number of opinion strings is $N_{op} = 2^J$. We denote by $Y^{(s)}$ the $s-th$ opinion string, $s = 1, \ldots, N_{op}$ where the opinion strings are indexed as $Y^{(1)} = (0, \ldots, 0), Y^{(2)} = (0, \ldots, 0, 1), \ldots, Y^{(N_{op})} = (1, \ldots, 1)$ so that $Y^{(s)}$ and $Y^{(N_{op}-s)}$ are opinion strings which disagree on every issue. We will also refer to opinion strings as opinion types. The opinion weights are given by $\{w_i\},\ i = 1, \ldots, J$. We shall assume that $\sum_{i=1}^{J} w_i = 1$. As in chapter 1, we define the weighted agreement, or similarity between two opinion strings s, s' to be the weighted average number of issues on which string s and string s' agree:

$$\sigma(s, s') = \sum_{i=1}^{J} w_i \cdot \delta(Y_i^{(s)}, Y_i^{(s')}),$$

where δ denotes the usual Kronecker delta. The threshold value α ($0 \leq \alpha \leq 1$) separates strings which are considered friends from those which are considered enemies so that opinion strings s, s' are considered friends if $\sigma(s, s') \geq \alpha$ and are enemies otherwise.

When a focus individual of type s meets with an interlocuter s' the focus individual's type is replaced with one of J (not necessarily distinct) types according to which issue is being discussed. We define the vector of change $F(s, s') = (s_1, \ldots, s_J)$ to be the J-dimensional vector such that s_i is the type that a focus individual of type s becomes after meeting with an individual of type s' concerning issue i. We assume that each issue is equally likely to be considered. To avoid ambiguity we will refer to the ith component of $F(s, s')$ as $F(s, s')_i$ rather than s_i. Note that F is determined by both α and the weights w_i.

For each opinion type s, let the proportion of individuals be denoted by x_s. Then we have

$$\sum_{s=1}^{N_{op}} x_s = 1. \tag{3.1}$$

The ODEs that govern their dynamics depend on the weights and the α threshold. We have

$$\dot{x_s} = \frac{1}{J(N-1)}\bigg(-\sum_{s'=1}^{N_{op}}\sum_{i=1}^{J} x_s x_{s'}[1-\delta(F(s,s')_i,s)] + \sum_{s'\neq s}\sum_{k=1}^{N_{op}}\sum_{i=1}^{J} x_{s'}x_k\delta(F(s',k)_i,s)\bigg), \tag{3.2}$$

for $1 \leq s \leq N_{op}$, where the first term in the large parentheses in equation (3.2) represents the proportion of the population of type s which becomes a different type and the second term represents the proportion of all other types which become type s.

Note that by (3.2),

$$\sum_{s=1}^{N_{op}} \dot{x_s} = 0, \tag{3.3}$$

that is, the total population of the community does not change.

Given J issues, the similarity function $\sigma(s,s')$ evaluated for all possible pairs of opinion types can take a maximum of N_{op} distinct values. This is because the value of $\sigma(s,s')$ is determined by which issues the two strings, s and s', have in common, as well as the weights on these issues. This maximum occurs if the weights are all different, and if, for any subset of weights $\{w_{n_i}\}_{i=1}^{m}$, we have the following conditions:

$$\sum_{i=1}^{m} w_{n_i} \neq w_p,\ p=1,\ldots,J \quad \text{and} \quad \sum_{i} w_{n_i} = \sum_{k} w_{m_k} \implies \{w_{n_i}\} = \{w_{m_k}\}. \tag{3.4}$$

That is, the sum of the weights in the subset is not equal to any of the weights, and if the sum is equal to the sum of any other subset of weights then the subsets are in fact equal. The maximum occurs in this case because there are $\binom{J}{j}$ ways in which s and s' can have

j issues in common. Under the conditions above, each of the $\binom{J}{j}$ ways gives a unique value for σ over all $j = 0, 1, \ldots, J$. Therefore we have

$$\sum_{j=0}^{J} \binom{J}{j} = 2^J$$

distinct values for the similarity function. The particular values depend on the weights. The minimum number of distinct values occurs when all of the weights are the same. In this case, the value of the similarity function is the same for each of the $\binom{J}{j}$ ways in which s and s' have j issues in common. This results in $J+1$ total values of the similarity function.

Suppose, for some choice of weights, there are a total of M distinct values of the similarity function. Let us arrange them in order from least to greatest:

$$0 = \sigma_1 < \sigma_2 < \cdots < \sigma_M = 1. \tag{3.5}$$

Here, $\sigma_1 = 0$ and $\sigma_M = 1$ correspond to zero and J issues being the same between the two individuals, respectively. The values of σ_k represent the thresholds in the set of α values. The behavior of the system (3.2) is identical for all values α, $\sigma_{k-1} < \alpha \leq \sigma_k$, $1 \leq k \leq M$. Therefore, beginning with $\alpha = 0$, the behavior of the system changes as α increases through each of the thresholds yielding M different model behaviors.

3.3 Solutions of the ODE's

3.3.1 The uniform solution

It can be shown that system (3.2) has a steady state,

$$x_s = \frac{1}{N_{op}}. \tag{3.6}$$

This steady state is neutrally stable for all values of α, that is, $\lambda = 0$ is an eigenvalue of the Jacobian corresponding to this uniform solution (3.6). The presence of a zero eigenvalue results from from the solution symmetry, equation (3.3), and corresponds to the eigenvector $(1, \ldots, 1)$. Let M_0 be the multiplicity of the zero eigenvalue. M_0 changes with α. If $M_0 = 1$ then from (almost) any initial condition satisfying (3.3) the system will converge to solution (3.6). In this case, every opinion is equally represented and any deviation from this solution will tend to decrease. In the stochastic equivalent of such a system will be characterized by oscillations around solution (3.6) and convergence to a one- or two-party final state will take a very long time.

If $M_0 > 1$, this indicates that, in the deterministic system, a neutral manifold of solutions is present. In the stochastic system of chapter 1, there will be neutral drift within this manifold, which eventually results in convergence to a fixed state. This happens on a much faster time-scale as compared to the situation described above where solution (3.6) is attracting.

In subsection 3.3.2, we study the behavior of the system (3.2) in the case that $J = 3$. Specifically we will derive solutions for $\alpha \in (0, \sigma_2]$ using an algorithm that can be generalized to any $J \geq 2$. We will also show that these solutions also satisfy the system when $\alpha = 0$. Finally we will demonstrate that the only stable solution the deterministic system has is (3.6) for α in the largest agreement score interval. In section 2.3 we will look at some cases for $J = 4$ issues. In particular, we will show that only solution (3.6) holds when $\delta = \pm 0.6$ which will explain the behavior of the stochastic model observed at these points in chapter 1 when the population is large.

3.3.2 Example: Solutions for $J = 3$ issues

In this section we assume throughout that $J = 3$. When examining α is in the largest agreement score interval we will assume that the weights are the same for notational

conveniences. While solution (3.6) is always an equilibrium, depending on the regime it belongs to different neutral solution manifolds.

For α in the interval $(0, \sigma_2]$ the following equalities characterize the equilibrium:

$$\begin{aligned} &x_1x_4 = x_2x_3, \quad x_1x_7 = x_3x_5, \quad x_1x_6 = x_2x_5, \\ &x_3x_8 = x_4x_7, \quad x_5x_8 = x_6x_7, \quad x_2x_8 = x_4x_6. \end{aligned} \tag{3.7}$$

To visualize this solution, create a "hypercube" network where the nodes are the $N_{op} = 2^J$ opinion types and the edges connect the types that are different by exactly one opinion, see Figure 3.1. In this network, each node is part of J square faces of the hypercube. The only pairs of enemies are those corresponding to pairs of opposing vertices. The particular weights on each issue play no role here. The expressions (3.7) state that for each of the J faces, the products of the opposite corners are equal to each other. Note that clearly, solution (3.6) is a subset of this solution. Also, it is easy to show that the following equations are a consequence of equations (3.7):

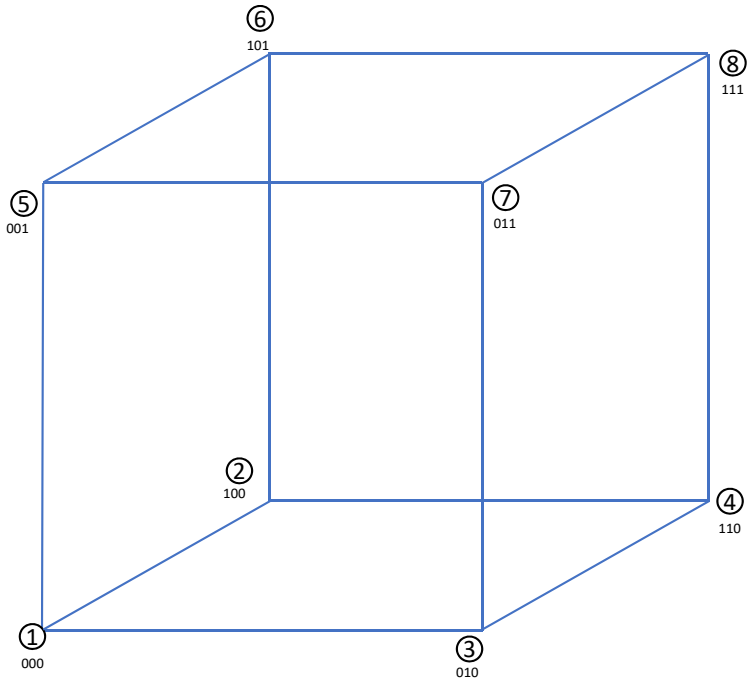


Figure 3.1: An example of a hypercube corresponding to $J = 3$, where each node is an opinion string and edges are drawn between nodes of Manhattan distance 1.

$$x_1x_8 = x_2x_7 = x_3x_6 = x_4x_5 \tag{3.8}$$

which state that the products of the opposite corners of the hypercube (that is, the types that are complete opposites of each other) are equal to each other.

Not all equalities in (3.7) are needed to uniquely define this solution, some are consequences of others. It is possible to show that the following equations imply all of equations (3.7):

$$x_2x_8 = x_4x_6, \quad x_1x_7 = x_3x_5, \quad x_2x_7 = x_4x_5 = x_3x_6. \tag{3.9}$$

The first two equalities describe two opposite 2D faces of the hypercube, and the last two are about pairs of opposite vertices. Figure 3.2 plots the products x_1x_4 and x_2x_3 in the stochastic model of chapter 1. Observe that these products tend to follow each other closely, just as we would expect from (3.7).

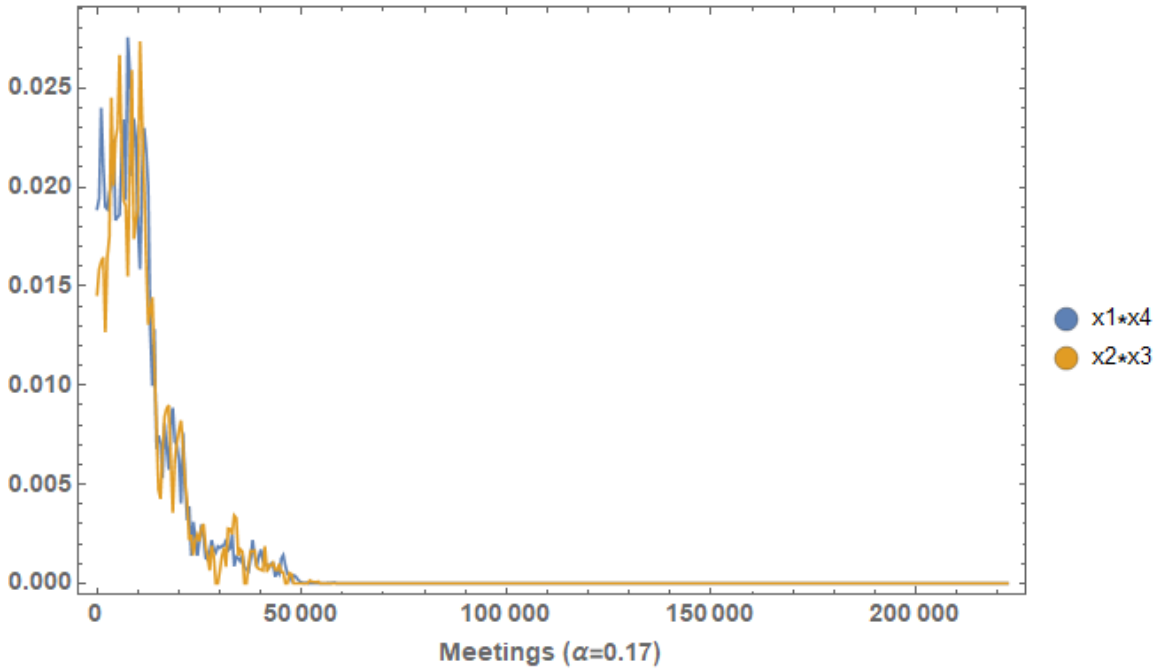


Figure 3.2: Plot of x_1x_4 and x_2x_3 in the stochastic model of chapter 1. Parameters are $N = 300,\ J = 3$, weight on each issue is $1/3$, and $\alpha = 0.17$.

We demonstrate a substitution algorithm to arrive at equations (3.9) in steps. Instead of using $x_1, x_2, \ldots, x_8$ it will be helpful to use the actual opinion strings for each opinion type, so, $x_{000}, x_{100}, \ldots, x_{111}$.

Step 1: Choose vertex 1 in figure 3.1 and note its opinion type (000).

Step 2: Note all the vertices (opinion types) which are Manhattan distance 2 from vertex 1 $(101, 011, 110)$.

Step 3: Formally add the strings corresponding to the opinion types of vertex one's neighbors $(100, 010, 001)$ to get each of the opinion types noted in Step 2. For example, $101 = 100 + 001$. Next, in system (3.2) make the formal substitution

$$x_{101} = \frac{x_{100} \cdot x_{001}}{x_{000}}$$

That is,

Step 4: For each opinion type s_i from step 2, if $n_j + n_k = s_i$ (where n_j, n_k are neighbors of 000) substitute

$$x_{s_i} = \frac{x_{n_j} \cdot x_{n_k}}{x_{000}}$$

into equations (3.2). Note that by only using neighbors of the vertex corresponding to type 000 the equation $n_j + n_k = s_i$ is unique.

Step 5: Now add opinion strings of neighbors of 000 to get all (one) opinion types Manhattan distance 3 from 000: $111 = 100 + 010 + 001$.

Step 6: From $111 = 100 + 010 + 001$ make the substitution

$$x_{111} = \frac{x_{100} \cdot x_{010} \cdot x_{001}}{x_{000}^2}$$

These substitutions solve system (3.2) for $\alpha \in (0, \sigma_2]$. Furthermore, these substitutions imply (after changing back to type numbers as indices)

$$x_1x_4 = x_2x_3,\ x_1x_6 = x_2x_5,\ x_1x_7 = x_3x_5,\ x_1x_8 = x_2x_7 = x_4x_5 \tag{3.10}$$

which are equivalent to (3.9). This can be generalized to other numbers of issues:

Let J be the number of issues. On a hypercube of dimension J if s_i is Manhattan distance $2 \leq d \leq J$ from the vertex corresponding to the string of all 0's (call this vertex 1) then there exists a unique set of neighbors of vertex 1 with corresponding strings $n_{k_1}, n_{k_2}, \ldots, n_{k_m}$ so that

$$s_i = \sum_{j=1}^{m} n_{k_j}.$$

Furthermore, the substitution

$$x_{s_i} = \frac{\prod_{j=1}^{m} x_{n_{k_j}}}{(x_1)^{m-1}} \tag{3.11}$$

for all vertices s_i Manhattan distance d ($2 \leq d \leq J$) from vertex 1, solves system (3.2) for $\alpha \in (0, \sigma_2]$.

Returning to $J = 3$, we now show that the solution (3.9) derived above also solves the system when $\alpha = 0$. Let $\dot{x}_s^{(0)}$ and $\dot{x}_s^{(0,\sigma_2]}$ be equations (3.2) for the two respective cases. Assume first that $\dot{x}_s^{(0,\sigma_2]} = 0$ for all s so that we have equalities (3.8). In the $\alpha = 0$ regime if type s and type s' disagree on all three issues (they lie on opposing vertices in figure 3.1) $\dot{x}_s^{(0)}$ will have a $-3x_s x_{s'}$ term in it since types s and s' are friends and the strings corresponding to these types differ in all three issues. It also follows, for each neighbor n of s, there will be an $x_n x_{n'}$ term. In fact, we have

$$\begin{aligned}
\dot{x}_1^{(0)} &= \dot{x}_1^{(0,\sigma_2]} - 3x_1x_8 + x_2x_7 + x_3x_6 + x_4x_5 \\
\dot{x}_8^{(0)} &= \dot{x}_8^{(0,\sigma_2]} - 3x_1x_8 + x_2x_7 + x_3x_6 + x_4x_5 \\
\dot{x}_2^{(0)} &= \dot{x}_2^{(0,\sigma_2]} + x_1x_8 - 3x_2x_7 + x_3x_6 + x_4x_5 \\
\dot{x}_7^{(0)} &= \dot{x}_7^{(0,\sigma_2]} + x_1x_8 - 3x_2x_7 + x_3x_6 + x_4x_5
\end{aligned} \tag{3.12}$$

and so on such that if s and s' are opposing opinion types $\dot{x}_s^{(0)}$ and $\dot{x}_{s'}^{(0)}$ have the same coefficients for the terms $x_1x_8,\ x_2x_7,\ x_3x_6\ x_4x_5$.

Consider $\dot{x}_1^{(0)}$. By equations (3.8), $-3x_1x_8 = -x_2x_7 - x_3x_6 - x_4x_5$ so $\dot{x}_s^{(0)} = 0$ By analogous arguments $\dot{x}_s^{(0)} = 0$ for all s.

However, the system $\dot{x}_s^{(0)}$ has other solutions. For example, setting $x_1, x_2, x_7, x_8 = 0.25$ provides a solution which violates (3.8).

Next we consider $\alpha \in (\sigma_3, 1]$, where we have assumed that the three weights are equal

so that there are four distinct values of σ. We will call this the *highest* α *regime*. Observe that in this regime, if t and t' are opposing types, we have

$$\dot{x}_t - \dot{x}_{t'} = -2(1 - x_t - x_{t'})(x_t - x_{t'}).$$

Therefore, if $\dot{x}_s = 0$ for each s then there are two cases. In the first case, $(1 - x_t - x_{t'}) = 0$ for some t so that $x_t + x_{t'} = 1$. Therefore $x_s = 0$ for each $s \neq t, t'$. However, this solution is unstable due to the Jacobian associated with this solution having a positive eigenvalue.

In the second case $(1 - x_t - x_{t'}) \neq 0$ for each t. Therefore the following holds:

$$x_1 = x_8,\ x_2 = x_7,\ x_3 = x_7,\ x_4 = x_5 \tag{3.13}$$

Substituting these into $\dot{x}_s$, $s = 1, \ldots 4$ and adding pairs of the resulting equations yield the following relations:

$$x_1 x_4 = x_2 x_3,\ x_1 x_3 = x_2 x_4,\ x_1 x_2 = x_3 x_4 \tag{3.14}$$

Now, each of x_i, $i = 1, \ldots 4$ is nonzero. Indeed, suppose e.g. $x_1 = 0$. Then by the first equation in (3.14) either $x_2 = 0$ or $x_3 = 0$. If $x_2 = 0$ then the third equation implies that $x_3 = 0$ or $x_4 = 0$. However, both cannot be zero by (3.1). Therefore, supposing $x_3 = 0$ and $x_4 \neq 0$ we must have $x_4 = x_5$ and all other population proportions are zero, contrary to our assumption that $(1 - x_t - x_{t'}) \neq 0$ for each t.

By making substitutions implied by (3.14) we have that $x_1 = x_2 = x_3 = x_4$ and by (3.13) it is seen that this is the uniform solution.

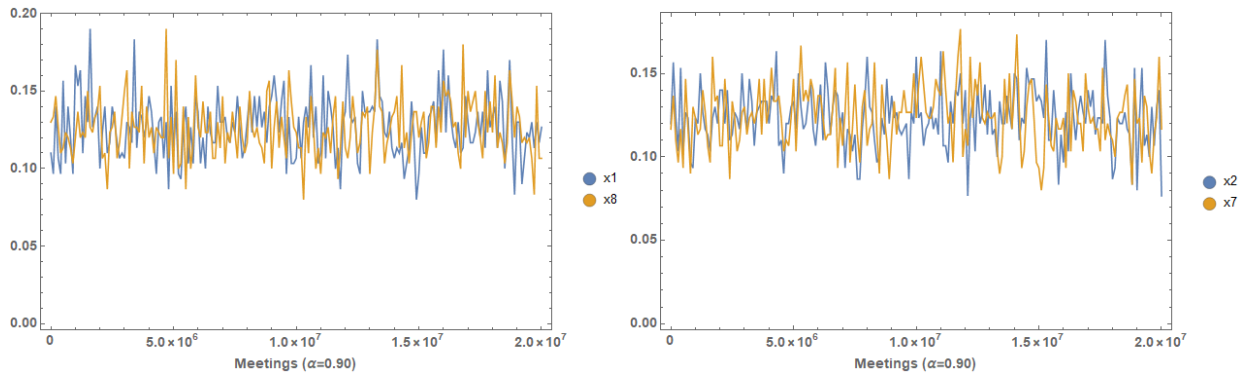


Figure 3.3: Plots of $x_1,\ x_8$ and $x_2,\ x_7$ in the stochastic model of chapter 1. Parameters are $N = 300,\ J = 3$, weight on each issue is $1/3$, and $\alpha = 0.90$.

In figure 3.3 we plot the behavior of four of the type proportions in the stochastic model for the highest alpha regime. This corresponds to a point for which $\delta = 1$ in chapter 1. We have paired together enemies $x_1,\ x_8$ in the left pane, and $x_2,\ x_7$ in the right pane. Observe the behavior is as expected - each of the type proportions oscillates about $1/8$, the uniform solution for the duration of the time that the simulation was allowed to run. The time it takes for the stochastic model to reach a fixed state will be, on average, extremely high.

3.3.3 Examples: $J = 4$ issues

We now turn to some examples when the community interacts with each other regarding four issues. For $\alpha \in (0, \sigma_2]$ the algorithm presented in section 3.3.2 yields the following relations to solve the system:

$$x_1x_4 = x_2x_3,\ x_1x_7 = x_3x_5,\ x_1x_6 = x_2x_5 \tag{3.15}$$

$$x_1x_{10} = x_2x_9,\ x_1x_{13} = x_5x_9,\ x_1x_{11} = x_5x_9$$

$$x_1x_8 = x_4x_5 = x_2x_7,\ x_1x_{12} = x_4x_9 = x_2x_{11} \tag{3.16}$$

$$x_1x_{14} = x_6x_9 = x_2x_{13},\ x_1x_{15} = x_7x_9 = x_3x_{13}$$

$$x_1x_{16} = x_8x_9 = x_5x_{12} = x_3x_{14} = x_2x_{15} \tag{3.17}$$

Referring to figure 3.4, (3.15) says that for each of the 6 faces to which vertex 1 belongs, products of diagonal vertices are equal; (3.16) says that for each of the 4 prisms vertex 1 belongs to, products of opposing vertices are equal; finally (3.17) says that products of vertices on the outer cube with the corresponding opposite vertices on the inner cube are equal.

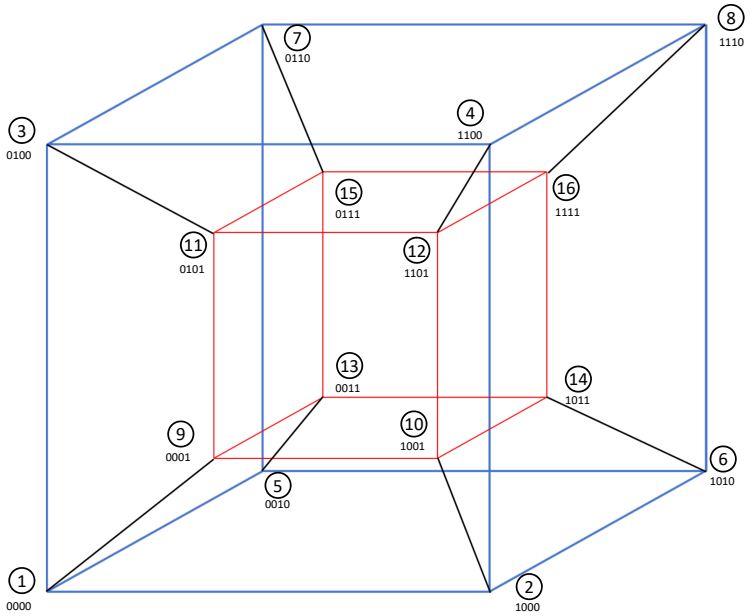


Figure 3.4: An example of a hypercube corresponding to $J = 4$, represented as a cube embedded in a cube. Each node is an opinion string and edges are drawn between nodes of Manhattan distance 1 as in figure 3.4. The blue and red edges connect vertices of the outer and inner cubes, respectively. The black edges connect vertices of the outer cube with vertices of the inner cube.

We also have a similar relationship between the $\alpha = 0$ and $\alpha \in (0, \sigma_2]$ regimes here, as

when $J = 3$. Since, for example, it can be shown that

$$\dot{x}_1^{(0)} = \dot{x}_1^{(0,\sigma_2]} - 4x_1x_{16} + x_2x_{15} + x_5x_{12} + x_8x_9$$

it can also be shown that the solution manifold for the system $\dot{x}_s^{(0,\sigma_2]}$ also solves $\dot{x}_s^{(0)}$; furthermore, just as with the $J = 3$ situation, so also does $\dot{x}_s^{(0)}$ have other solutions.

Also, the largest α regime has only the uniform solution as a stable solution. However the method for deriving this conclusion in $J = 3$ does not have an obvious analog here. That is to say,

$$\dot{x}_1 - \dot{x}_{16} \neq C(x_1 - x_{16})(1 - x_1 - x_{16})$$

as would be expected if the method from $J = 3$ were to translate here. However, the multiplicity of the eigenvalue $\lambda = 0$ of the Jacobian corresponding to the uniform solution is $M_0 = 1$ indicating that (3.6) is the only stable solution.

This is also the case with parameters corresponding to $\delta = \pm 0.6$. As we saw in chapter 2 the number of meetings to reach a fixed state increased rapidly at higher populations. The presence of the uniform solution as the only stable one explains why this is the case, see figure 3.5.

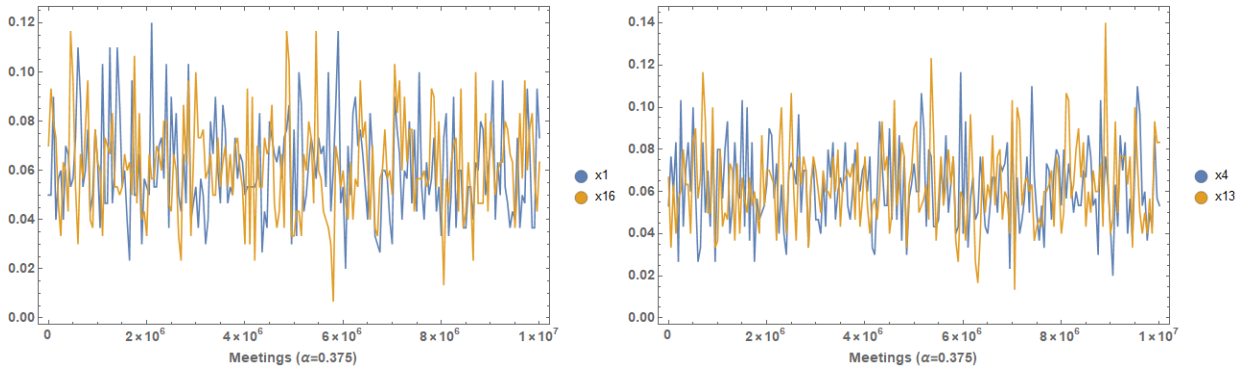


Figure 3.5: Plots of $x_1,\ x_{16}$ and $x_4,\ x_{13}$ in the stochastic model of chapter 1. Parameters are $N = 300,\ J = 4$, weight on each issue is $1/4$, and $\alpha = 0.375$.

3.4 Conclusion

In this chapter we developed a system of ODEs as an analog to the stochastic model studied in chapter 2. The main result here was the discovery of a neutrally stable solution that exists in all cases such that when this uniform solution is the only stable solution, the opinion type population proportions in the stochastic model will oscillate about this solution. This behavior explains the extremely long times to fixation seen in certain situations. If δ is close to -1, there is an intuitive explanation for this phenomenon. However, the result presented here gives an explanation which is derived from the mathematical properties of the underlying deterministic system. Such an explanation is helpful when the same phenomenon occurs ($\delta = \pm 0.6$ at larger populations in the stochastic system), but there is no clear, intuitive explanation based on the switching properties for these values of δ.

We also presented an algorithm which solves the system of ODEs for any number of $J \geq 2$ issues when $\alpha \in (0, \sigma_2]$ and demonstrated that these solutions also solve the system corresponding to $\alpha = 0$ when $J = 3, 4$.

Chapter 4

The structure of the HPV vaccine re-tweet network

4.1 Introduction

Major problems that are faced by society today include protecting our health, and making sure that there is enough food, water, and energy on the planet for us, our children, and future generations. The rise of social media platforms such as twitter provide a forum where these problems are increasingly discussed by members of society - to such an extent that individual, and collective opinions are greatly influenced by those discussions. Public opinion and individual beliefs are crucial for our ability, as a society to solve these crucial challenges. And a problem that is detrimental to the process is the spread of misinformation through social media.

For example, the spread of misinformation counters our ability to limit and control infectious diseases, e.g. through preventing the effective implementation of vaccination (given the doubt about vaccination that is being propagated). Similarly, misinformation about the

human role in climate change is harmful for any measure that aims at limiting the negative impact of humans.

The dynamics of information and misinformation spread across social networks is currently not well-understood. Improving our understanding in this respect and designing effective intervention measures to prevent the spread of misinformation are vital for devising strategies that shift the dynamics in favor of the spread of true information. It is also a necessary step in any attempt to solve the large-scale problems faced by the society.

Mathematical/computational approaches that aim to predict the spread of true and false information and that can potentially suggest intervention strategies to advance the adoption of true, scientific information by populations, have so far been largely missing. At the same time, mathematics can offer a solution to the problem of misinformation spread. There are strong parallels between mathematical descriptions of information spread through communities, and the mathematical descriptions of infectious disease spread from host to host. Information, or ideas, spread like memes, from the mind of one individual to the next, through cultural transmission processes. In this process, different types of information (e.g. true vs. false) can compete for susceptible minds, similar to the competition of two pathogen strains for susceptible hosts. Here we build data-driven mathematical models of information spread to fill this gap. This is done specifically in the context of HPV vaccination, and study the spread of information through Twitter. The basic idea is to harvest social media data to build networks, which will inform mathematical models of information spread across these networks. Methodologies used here can be applied to further studies of the competition between information and misinformation, and the identification of individuals that are “on the fence” and can thus be more effectively targeted by intervention approaches.

4.2 Data set and network definition

We begin with twitter meta data obtained from October 6, 2018 through November 28, 2019. These data were extracted from the Twitter Filtered Stream API v1.1, and Twitter Retweet API v1.1 using keywords such as “hpv vaccine”, and “hpvvaccination”. Tweets and retweets were extracted which contained any of the keywords, including those where a keyword was a part of a “hashtag”. The data set contains $12,171$ tweets concerning the HPV vaccine which were re-tweeted at least once. Included in the data is an encoded user identification number of the original tweeter, the date and time of the original tweet, the content of the original tweet, the number of times it was re-tweeted, the identification number of the user that re-tweeted, and finally the date and time that the re-tweet occurred. Each row in the data set corresponds to a unique re-tweet of a tweet so that if a particular tweet were re-tweeted n times, there are n rows in which that tweet is the original tweet.

In order to analyze this data we define a network G whose nodes are users contained in the data set, and if $x,\ y$ are two nodes, there is an edge from x to y if y re-tweeted a tweet by x signifying that information has traveled from x to y. We can denote such an edge as an ordered pair (x, y) since information is traveling in one direction. By this definition, the edges of G are naturally directed. However, since our network is constructed out of data which occurred in time, and since community detection within networks is typically conducted by either assuming, or forcing G to be undirected [15] we construct the network so that if there is an edge (x, y) or (y, x) we regard it to be an undirected edge, signifying that there is an established information conduit between x and y irrespective of the direction it occurred in the past. This network G is called the *re-tweet network*, see figure 4.1.

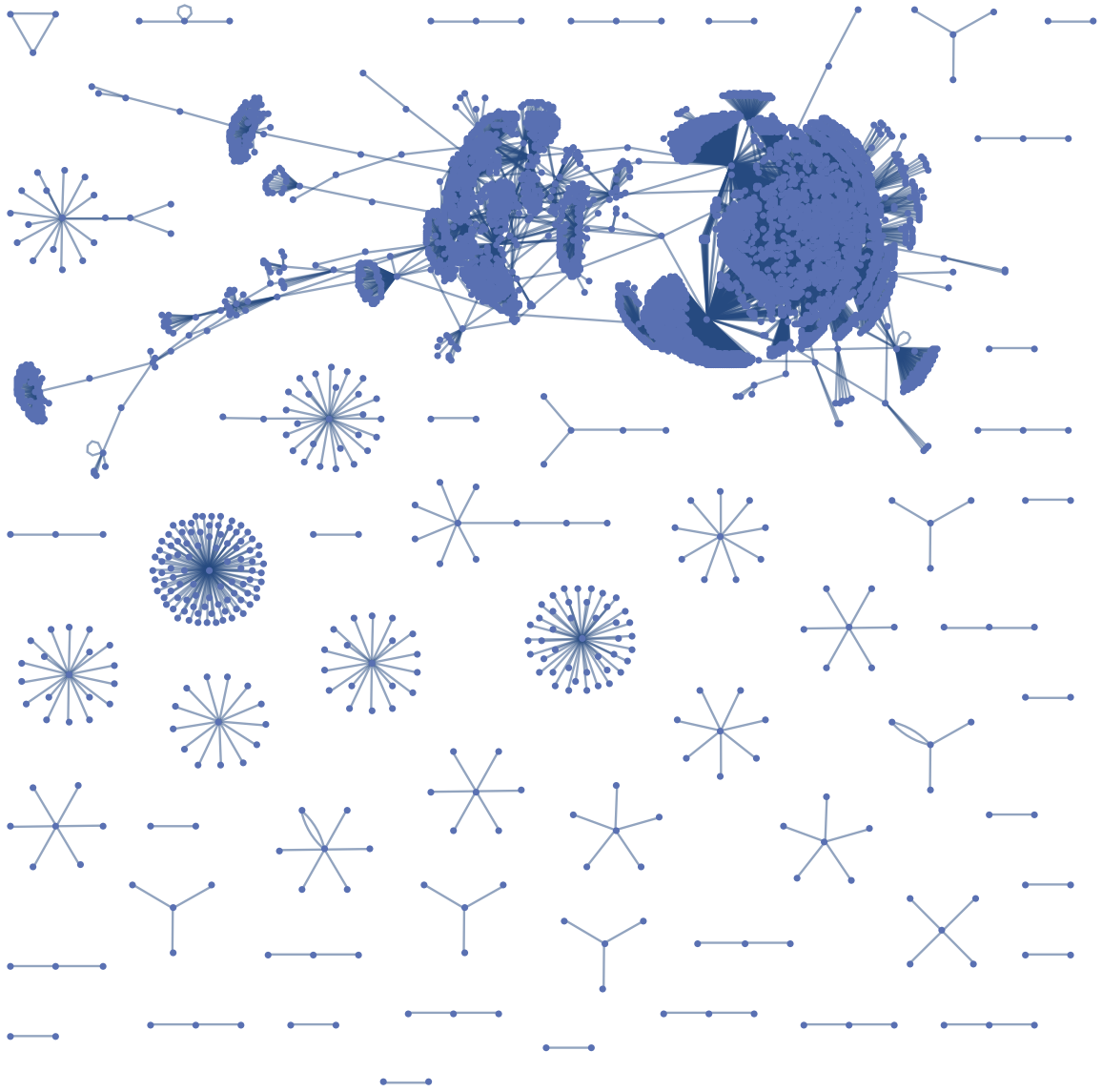
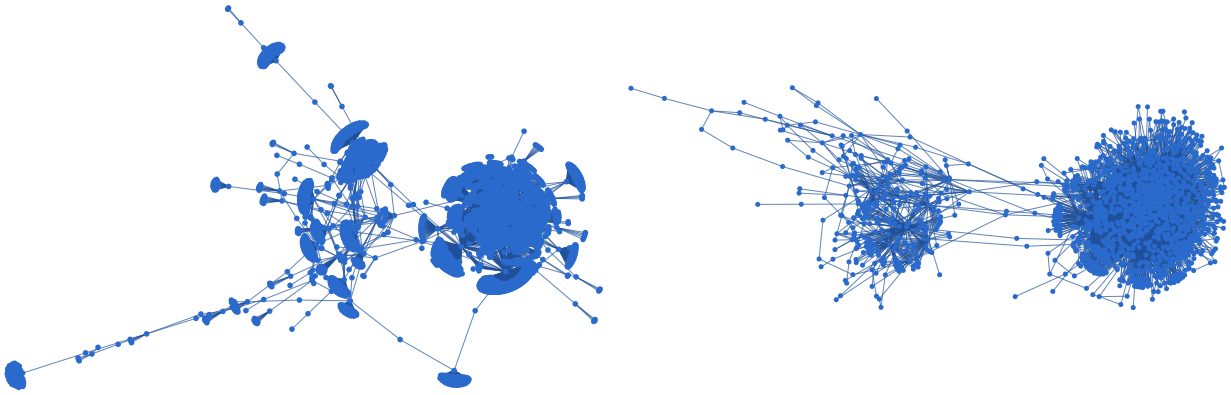


Figure 4.1: The constructed re-tweet network G.

We now further reduce the size of the network under analysis. First, observe that there is one large connected component and dozens of much smaller connected components. We focus on the underlying simple graph of the large connected component (see figure 4.2a). Furthermore, note the many 'flower' shapes in figure 4.2a. These represent users who either (1) only re-tweeted one user's tweet and were not themselves re-tweeted; or (2) a multitude of users which one and only one user re-tweeted. In either case, the underlying



(a) With ”flower” shapes (b) Without ”flower” shapes

Figure 4.2: The largest connected component of G. (a) before the “flower” shapes are removed; (b) after the flower shapes are removed.

structure of the network is not affected by removing these ‘flower’ shapes. The resulting network, G', depicted in figure 4.2b, is composed of 1409 nodes, and 4185 edges as opposed to 5304 nodes and 7169 edges in the network which includes the ‘flower’ shapes. figure 4.2a.

4.3 Finding communities in the network

The network G' shown in figure 4.2b was drawn by using software *Mathematica*, which implements a force-directed graph drawing algorithm to present graphs in the most aesthetically pleasing way, with minimum edge intersection. This visually suggests the existence of two sub-communities, corresponding to a loosely connected left side and a more densely connected right side. In what follows we (1) apply a systematic algorithm to assign all nodes to one of two communities and (2) determine if these communities are meaningful in the context of the tweets which the members of each community represents.

In this section we use a label propagation algorithm developed by Raghavan, Albert, and Kumara [20] to answer the first of these questions. Specifically, each of these methods find distinct communities within the network. Afterwards, unions of these communities then form the left and right communities mentioned above. The investigation of the second question is visited in section 4.4.

The label propagation algorithm we use is based on the following idea [20]: if a node x in the network G' has neighbors $x_1, x_2, \ldots, x_k$ and the nodes of G' have a label denoting a community to which each belongs, then the label of x is determined based on its neighbors. Specifically, we initialize the algorithm by assigning each node in the network a unique label. Then, in random sequential order, we select a node x and, if possible, the label of x is updated to be the label carried by the majority of its neighbors, with ties broken uniformly randomly. This process is performed iteratively, where at each iteration, the order in which the nodes update their labels according to their neighbors is chosen randomly.

If $l_1, \ldots, l_p$ are the labels which currently exist in the network and $n_i^{(l_m)}$ is the number of neighbors of node i with label l_m then the algorithm stops if for each node i,

$$\text{node } i \text{ has label } l_m \text{ then } n_i^{(l_m)} \geq n_i^{(l_j)} \quad 1 \leq j \leq p$$

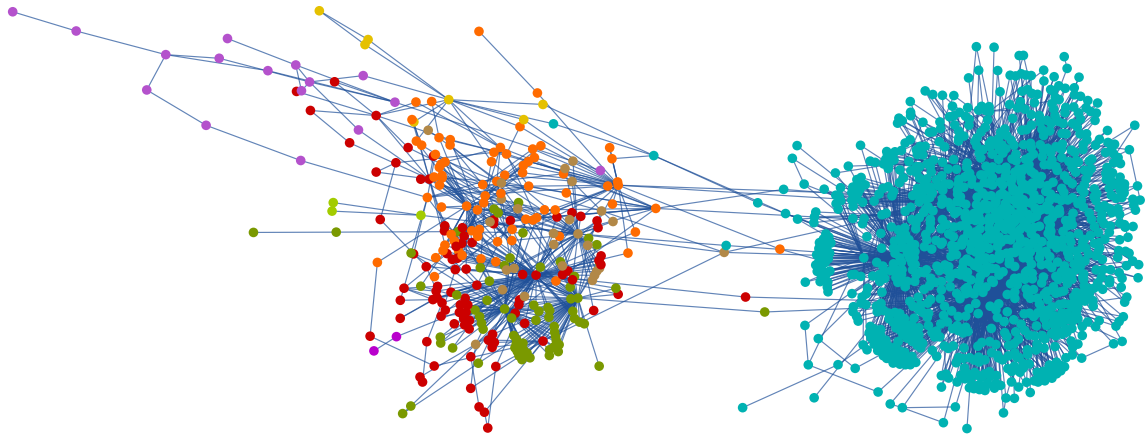


Figure 4.3: Results of label propagation algorithm. Different colored nodes depict communities defined by the algorithm.

The communities are then the collections of nodes with identical labels. Of course, due to the random initialization of the node labels, the random ordering of the nodes at each iteration, and the random breaking of ties specific results may differ in each implementation. However, figure 4.3 displays the typical result of the propagation algorithm. Observe that the algorithm identifies nine sub-communities. The largest subcommunity (of 1123 nodes) coincides with the dense right hand side of the network, and the rest of the communities (containing the total of 286 nodes) comprise the left hand side. In other implementations of the label propagation algorithm different numbers of subcommunities were found, but the algorithm always identified a single community associated with the densely connect right side of G' and then smaller communities on the left side. Additionally, the middle nodes which join the two sides have been placed into either the right side community or into one of the communities on the left side. We define the 'Right'community to be the nodes with the cyan color and the 'Left'community to be the union of the nodes of all other colors, see figure 4.4.

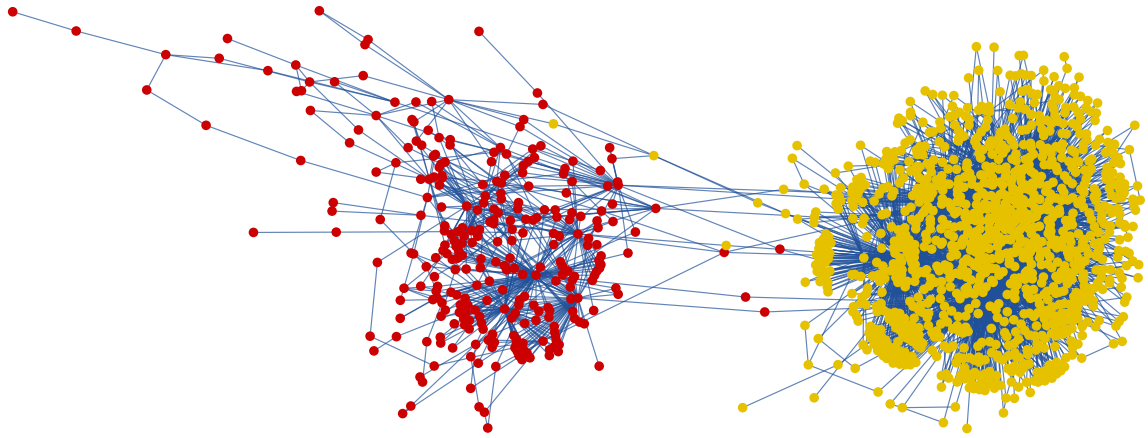


Figure 4.4: The Left (red) and Right (yellow) communities of nodes

4.4 Analysis of tweet contents in sub-components

We studied the results of the label propagation algorithm, to determine if the Left and Right communities defined in section 3 were meaningful with respect to the language of the tweets which were retweeted. To this end, we used volunteers to read and classify a subset of tweets from the Left and Right communities. Then Cultural Consensus methodology [23] was employed to analyze the responses and classify the tweets. The details are presented below.

4.4.1 Survey data and initial analysis

The ultimate goal of this analysis is to determine if the tweets re-tweeted by individuals from the Left and Right communities are systematically different. To identify possible differences, we took two random samples of 100 nodes from the network on figure 4.4 such that 50 of the nodes are in the Left community and 50 of the nodes are in the right community. The content of each of the 100 tweets, in a randomized order, was then sent to a number of volunteers. These individuals were asked label each tweet as P, A, or U

depending on whether they thought the tweet was pro-HPV vaccine, anti-HPV vaccine, or undecided, respectively. 15 individuals recorded their responses for the first survey and 16 individuals recorded responses for the second survey. With the exception of one person, everyone who participated in the first survey also participated in the second. Additionally, all but two participants in the second survey also participated in the first survey. In order to avoid population differences between the survey participants, we did not use any data from participants who only took part in one survey. Thus our analyzed survey data consists of the results from the 14 individuals who took both surveys. These 14 participants were comprised of seven men and seven women. Their ages ranged from 18 to 66. The highest education level attained amongst these individuals is a doctoral degree; the lowest is high school graduate.

Responses of the participants to the first survey are summarized in figure 4.5, and responses to the second survey in figure 4.6. This information is further presented in figures 4.8 and 4.8 as histograms, which show, for each survey, the distributions of participant responses that classified the tweets as A (green), P(orange), or U (blue). The left panels (marked as (a)) show data for the tweets that came from the Left community, and the right panels (marked (b)) show tweets from the Right community. It is clear that for both surveys, the majority of tweets that came from the Left community were categorized as A, and the majority of tweets from the Right community were categorized as P. This gave rise to the hypothesis that the Left community contains tweets that are against the HPV vaccine, while the right community contains tweets that are pro-vaccine. Therefore, a key was created that associated all the tweets from the Left community with response A, and all the tweets from the right community with response P.

On the first survey, participants' responses matched the key a minimum of 43 times and a maximum of 94 times with a mean of 76.4 matches. On the second survey the minimum number of matches was 39, the maximum was 95, and the mean number of matches was

Survey response distribution (Key = A)

Participant	P	U	A
1	0	17	33
2	4	18	28
3	1	10	39
4	1	16	33
5	2	0	48
6	3	2	45
7	5	19	26
8	4	4	42
9	1	9	40
10	3	4	43
11	8	0	42
12	1	10	39
13	2	18	30
14	0	13	37
Mean	2.5	10	37.5

(a) Tweets from the Left community, Survey 1.

Survey response distribution (Key = P)

Participant	P	U	A
1	33	17	0
2	37	11	2
3	38	10	2
4	34	14	2
5	46	3	1
6	44	2	4
7	31	10	9
8	42	5	3
9	43	5	2
10	39	9	2
11	43	0	7
12	43	6	1
13	32	17	1
14	39	11	0
Mean	38.85714	8.571429	2.571429

(b) Tweets from the Right community, Survey 1.

Figure 4.5: Participant response distributions for survey 1. (a) Survey items for which the key = A; (b) items for which the key = P.

81.1. In performing a paired t-test these means are statistically different at a significance level of $\alpha = .10$ ($p = 0.083$). This suggests that, perhaps, the first survey was 'more difficult 'than the second. This is consistent with the number of U responses that were obtained: the participants, on average, found more items ambiguous or undecided (response U) from the left community than the right, and in survey 1 than in survey 2, see figures 4.7 and 4.8, blue bars. For both surveys, the mean number of matches (76.4 and 81.1 on the two surveys, respectively) suggest that survey participants tend to agree that the Left and Right communities separate tweets which are pro-HPV vaccine from those who are anti-HPV vaccine.

4.4.2 Cultural consensus construction

In order to strengthen the evidence that the communities illustrated in figure 4.4 consist of those who are pro and anti HPV vaccine, we use a data analysis method created by

Distribution of answers (key = A)

Participant	P	U	A
1	0	7	43
2	2	1	47
3	2	5	43
4	3	4	43
5	1	7	42
6	1	1	48
7	3	27	20
8	6	3	41
9	2	5	43
10	3	15	32
11	18	13	19
12	2	5	43
13	3	8	39
14	0	14	36
Mean	3.285714	8.214286	38.5

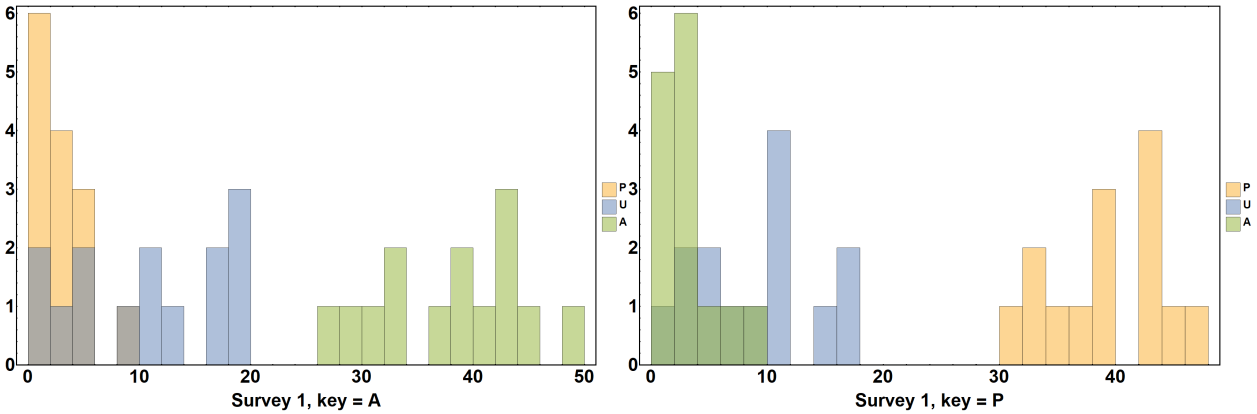
(a) Tweets from the Left community, Survey 2.

Distribution of answers (key = P)

Participant	P	U	A
1	48	2	0
2	48	0	2
3	45	4	1
4	43	2	5
5	45	5	0
6	47	1	2
7	35	12	3
8	45	3	2
9	48	1	1
10	39	9	2
11	20	10	20
12	45	3	2
13	42	7	1
14	47	3	0
Mean	42.64286	4.428571	2.928571

(b) Tweets from the Right community, Survey 2.

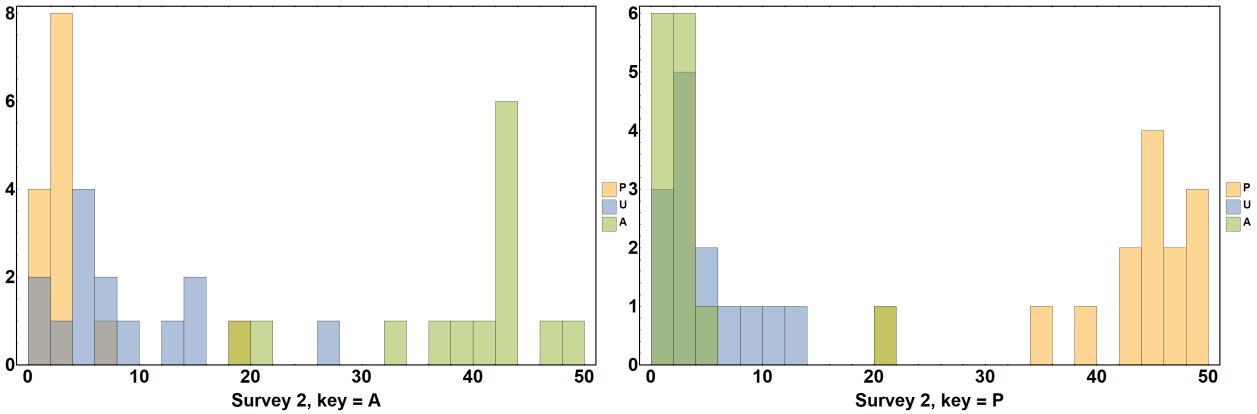
Figure 4.6: Participant response distributions for survey 2. (a) Survey items for which the key = A; (b) items for which the key = P.



(a) Key = A

(b) Key = P

Figure 4.7: Histograms of survey 1 responses (a) items from left community (i.e. key = A); (b) items from right community (key = P).



(a) Key = A (b) Key = P

Figure 4.8: Histograms of survey 2 responses (a) items from left community (i.e. key = A); (b) items from right community (key = P).

Romney, Weller, and Batchelder [23]. This methodology has been created to extract "cultural consensus", which allows for a more accurate knowledge representation compared to more standard methods such as majority counts. The cultural consensus method is especially useful when some of the questions are difficult or ambiguous, and when the level of participants "expertise" is variable.

The basic idea is to use the survey responses to create a new response consisting of likely correct answers. We can then compare this new list of responses to the key.

Before summarizing this method, we list the assumptions used in this analytical method and simultaneously introduce some notation.

1. *Common truth.* Each item k on a survey has a correct answer, Z_k. That is, there is a fixed answer key which is applicable to all survey participants.

2. *Local Independence.* That is, we assume that each respondent's answers are given independently of each other respondent. If $X_{i,k}$ is respondent i's response to item k, then mathematically the assumption is that the informant-item response random

variables satisfy condition independence

$$Pr[(X_{i,k})_{N\times M}|(Z_k)_{1\times M}] = \prod_{i=1}^{N}\prod_{k=1}^{M} Pr(X_{i,k}|Z_k)$$

where N is the number of respondents, M is the number of items, and as before, Z_k is the correct answer to item k.

3. *Homogeneity of the Items.* The assumption here is that each respondent i has a fixed 'cultural competence', D_i, over all survey items.

In this notation, our survey data yields two 14×100 matrices $(X_{i,k})_{14\times 100}$ (one for each survey) composed of the values $1,\ -1,\ 0$ where these values indicate an answer to an item is P,A, or U, respectively. We then infer the cultural competencies of each of respondents from the proportion of matches among them whence we infer the list of correct answers Z_k using Bayes theorem. In what follows, we summarize how this is done. For details refer to [23] and Comrey [3].

Recall from above that D_i is the fixed cultural competence of respondent i over all items. Define the random variables

$$M_{ij,k} = \begin{cases} 1 \text{ if } i \text{ and } j \text{ match on question } k \\ 0 \text{ otherwise,} \end{cases}$$

and let M_{ij} be the observed proportion of matches between respondents i and j over all items. Then $Pr(M_{ij,k}) = D_iD_j + [1 - D_iD_j]/L$ where L is the number of possible responses to each question. In our case $L = 3$. Since the right hand side is independent of k we can replace the left hand side with M_{ij}. As the M_{ij} are derived from survey data, we may

solve for D_iD_j to obtain estimates for these products and define

$$M^*_{ij} = D_i\tilde{D}_j = (LM_{ij} - 1)/(L - 1).$$

From this set of equations we may estimate the components of the factor vector $< D_i >_{i=1}^{14}$ using the iterative method in [3]. It should be noted that so long as assumptions 1,2,3 are reasonably satisfied, there is only one relevant factor that need be extracted [23], in this case, the competencies with which each respondent categorizes the tweets as P, A, or U. After getting estimates for each respondents cultural competency they are used along with Bayes' Theorem to compute the conditional probabilities

$$Pr(Z_k = l| < X_{i,k} >_{i=1}^{14})$$

for each possible $l = 1, \ldots, 3$ and each $k = 1, \ldots, 100$, where $< X_{i,k} >_{i=1}^{14}$ denotes the column vector of responses by all respondents to item k. It is here that the estimates for the D_i are used (the local independence assumption is also used here.) [23]. Choosing $Z_k = l$ for the l which maximizes the equation above yields the list of derived 'correct answers.'

4.4.3 Comparison between key and derived correct answers

Now that we have our list of correct answers we can compare them to the survey keys. The proportion of items on which the correct answers for survey 1 match the key is 0.86 and that same proportion for survey 2 is 0.93 indicating that there is strong agreement between the derived correct answers and the keys for both surveys, especially considering that some of the derived correct answers based on our survey data turned out to be U and there is no such answer in the keys. Indeed, the derived correct answers include 13 U's on survey 1 and 5 U's on survey 2. That is, out of the 14 non-matching answers on survey 1 all but

one came from an answer given as U. On survey 2, all but 2 of the non-matching answers came from answers given as U.

The statistic which is typically used in the social sciences to measure the level of agreement between a list coded annotations and that of an objective judge is Cohen's Kappa ([9], [10])

$$\kappa = \frac{\pi_o - \pi_e}{1 - \pi_e}$$

where π_e and π_o are, respectively, the expected and observed agreement frequencies between the two lists. In words, κ measures the rate of agreement after removing from consideration the expected rate of agreement. The annotation of Left and Right communities in section 2 is the list of coded annotations and the derived correct answers form the list of annotations from an objective judge.

We report $\kappa = 0.752$ and $\kappa = 0.867$ for survey 1 and 2 respectively. These fall in the "substantial" and "nearly perfect" agreement range [13]. However, when considering only those tweets which the derived correct answers marked as P or A and ignore those marked U we report $\kappa = 0.989$ for survey 1 based on 87 tweets, and $\kappa = 0.962$ for survey 2 based on 95 tweets, both of which fall in the "nearly perfect"range. Therefore we conclude that the survey response data show that the results of the propagation algorithm do a remarkably good job of placing tweet nodes in both the 'pro' and 'anti' HPV vaccine categories.

4.5 Conclusions

Our analysis shows that methods of network analysis can be used to study community opinions. In particular, based purely on network structure, we used a network community detection method (label propagation) to identify two subcommunities of twitters (Left and Right). Furthermore, we were able to show that these two subcommunities are clearly

meaningful with regards to the tweet contents associated with which node. The survey response data we obtained suggested that the Left (resp. Right) side of the network could be associated with anti- (resp. pro-) HPV vaccine twitters. Using the survey response data, we then constructed a cultural consensus of likely correct answers to each survey item. These corresponded nearly perfectly with anti- and pro- HPV vaccine twitters.

Bibliography

[1] R. Axelrod. The dissemination of culture: A model with local convergence and global polarization. *Journal of conflict resolution*, 41(2):203–226, 1997.

[2] C. Castellano, M. Marsili, and A. Vespignani. Nonequilibrium phase transition in a model for social influence. *Physical Review Letters*, 85(16):3536, 2000.

[3] A. L. Comrey. The minimum residual method of factor analysis. *Psychological Reports*, 11(1):15–18, 1962.

[4] J. C. González-Avella, M. G. Cosenza, V. M. Eguíluz, and M. San Miguel. Spontaneous ordering against an external field in non-equilibrium systems. *New Journal of Physics*, 12(1):013010, 2010.

[5] J. C. González-Avella, M. G. Cosenza, and M. San Miguel. A model for cross-cultural reciprocal interactions through mass media. *PLoS one*, 7(12):e51035, 2012.

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

[7] R. Hegselmann and U. Krause. Opinion dynamics driven by various ways of averaging. *Computational Economics*, 25(4):381–405, 2005.

[8] D. Kempe, J. Kleinberg, S. Oren, and A. Slivkins. Selection and influence in cultural dynamics. *Network Science*, 4(1):1–27, 2016.

[9] R. H. Kolbe and M. S. Burnett. Content-analysis research: An examination of applications with directives for improving research reliability and objectivity. *Journal of consumer research*, 18(2):243–250, 1991.

[10] K. Krippendorff. *Content analysis: An introduction to its methodology*. Sage publications, 2018.

[11] N. Lanchier et al. The Axelrod model for the dissemination of culture revisited. *The Annals of Applied Probability*, 22(2):860–880, 2012.

[12] N. Lanchier, S. Scarlatos, et al. Fixation in the one-dimensional Axelrod model. *The Annals of Applied Probability*, 23(6):2538–2559, 2013.

[13] J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174, 1977.

[14] K. I. Mazzitello, J. Candia, and V. Dossetti. Effects of mass media and cultural drift in a model for social influence. *International Journal of Modern Physics C*, 18(09):1475–1482, 2007.

[15] M. Newman. *Networks*. Oxford university press, 2018.

[16] L. R. Peres and J. F. Fontanari. The mass media destabilizes the cultural homogenous regime in Axelrod's model. *Journal of Physics A: Mathematical and Theoretical*, 43(5):055003, 2010.

[17] L. R. Peres and J. F. Fontanari. The media effect in Axelrod's model explained. *EPL (Europhysics Letters)*, 96(3):38004, 2011.

[18] A. Pluchino, V. Latora, and A. Rapisarda. Compromise and synchronization in opinion dynamics. *The European Physical Journal B-Condensed Matter and Complex Systems*, 50(1-2):169–176, 2006.

[19] A. Radillo-Díaz, L. A. Pérez, and M. del Castillo-Mussot. Axelrod models of social influence with cultural repulsion. *Physical Review E*, 80(6):066107, 2009.

[20] U. N. Raghavan, R. Albert, and S. Kumara. Near linear time algorithm to detect community structures in large-scale networks. *Physical review E*, 76(3):036106, 2007.

[21] A. H. Rodríguez, M. del Castillo-Mussot, and G. J. Vázquez. Induced cultural globalization by an external vector field in an enhanced Axelrod model. In *7th International Conference on Practical Applications of Agents and Multi-Agent Systems (PAAMS 2009)*, pages 310–318. Springer, 2009.

[22] A. H. Rodríguez and Y. Moreno. Effects of mass media action on the Axelrod model with social influence. *Physical Review E*, 82(1):016111, 2010.

[23] A. K. Romney, S. C. Weller, and W. H. Batchelder. Culture as consensus: A theory of culture and informant accuracy. *American anthropologist*, 88(2):313–338, 1986.

[24] P. Singh, S. Sreenivasan, B. K. Szymanski, and G. Korniss. Accelerating consensus on coevolving networks: The effect of committed individuals. *Physical Review E*, 85(4):046104, 2012.

[25] A. Sîrbu, V. Loreto, V. D. Servedio, and F. Tria. Opinion dynamics: models, extensions and external effects. In *Participatory sensing, opinions and collective awareness*, pages 363–401. Springer, 2017.

Appendix A

Agreement-disagreement score: additional calculations

In order to observe convergence time, two faction proportion, and mean largest faction size, with respect to different values of δ we need to count how many regions there are and then to pick a point in each region. We do this by associating points in the space $S = (0, \infty)^{J-1} \times (0, 1)$ with the 2^{J-1} $(J-1)$-long sequences with terms in $\{0, 1\}$. Let $\{a_i\}$ be the collection of all $J-1$ agreement curves. Choose any point $\bar{x} \in S$. Notice that for each i, $\bar{x} = (X_1, \ldots, X_{J-1}, \alpha)$ is either "above" or "below" a_i in the following sense: $a_i \geq \alpha$ or $a_i < \alpha$. In the former case we associate $\bar{x}$ with a binary sequence whose i-th term is 1 and in the latter 0. Thus every point $\bar{x} \in S$ corresponds to a $(J-1)$-long binary sequence. It is clear that this correspondence is surjective and that two points $\bar{x}$, $\bar{y}$ in S belong to the same region if and only if they correspond to the same sequence. In the case at hand (weights $\omega_1, \omega_2, \omega_3$ vary, while $\omega_4 = 1$ is fixed) 148 regions were detected by repeatedly sampling points from S as described above. Although 148 regions were found by our algorithm, the various regions correspond to 22 different values of δ.

	Issue 1		Issue 2		Issue 3		Issue 4	
Region	Type 1	Type 2	Type 1	Type 2	Type 1	Type 2	Type 1	Type 2
1	7	0	7	0	7	0	7	0
2	7	1	6	0	6	0	6	0
3	4	0	5	1	5	1	5	1
4	4	1	4	1	4	1	4	1
5	1	0	4	3	4	3	4	3
6	4	4	2	2	2	2	2	2
7	1	1	3	3	3	3	3	3
8	0	0	4	4	4	4	4	4
9	0	1	3	4	3	4	3	4
10	1	4	1	4	1	4	1	4
11	1	7	0	6	0	6	0	6
12	0	4	1	5	1	5	1	5
13	0	7	0	7	0	7	0	7

Table A.1: The switch configuration of each region - one weighted issue

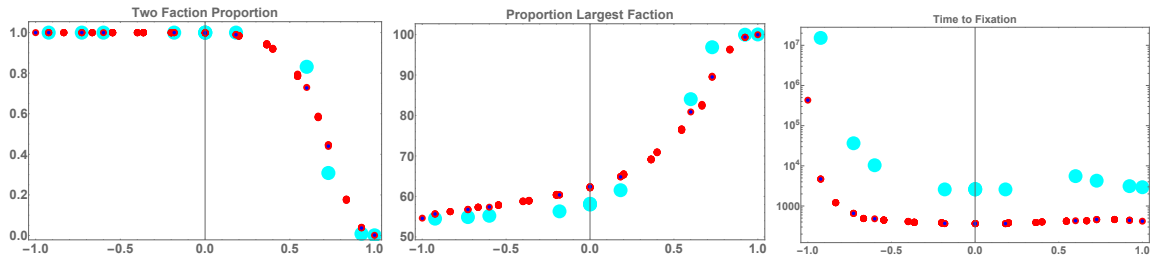


Figure A.1: Same as figure 2.4, but showing results for a larger population size ($N = 25$). Plotted are: the two faction proportion (a), mean largest faction size (b), and the mean time to fixation (c), plotted against δ. Large cyan points correspond to the simulations with $N = 25$ and one weighted issue. As in figure 2.4, blue (red) points correspond to the $N = 10$ system with a single weighted issue (all weighted issues). Each point represents the mean over 20,000 independent simulations. Standard error is too small to be visible on the graph.

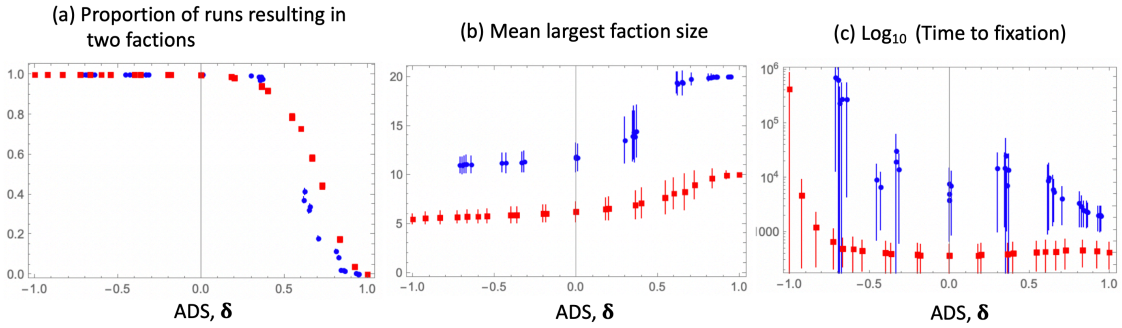


Figure A.2: Individual weighting of the issues, larger population size. Same as figure 2.5, but showing results for simulations with $N = 20$, were each person assigned weights individually (blue). This is compared with $N = 10$ results with uniform weighting (red). Plotted are: the two faction proportion (a), mean largest faction size (b), and the mean time to fixation (c), plotted against δ. As in figure 2.4, red points correspond to weights being the same for all individuals in the population (all weighted issues). For $N = 20$; each point represents the mean over 1,000 independent simulations. Vertical bars are standard deviations.