Persistence of Anti-vaccine Sentiment in Social Networks Through Strategic Interactions

A S M Ahsan-Ul Haque[1,2], **Mugdha Thakur**[1], **Matthew Bielskas**[1, 2], **Achla Marathe**[1,3], **Anil Vullikanti**[1,2]

[1] Biocomplexity Institute, University of Virginia
[2] Department of Computer Science, University of Virginia
[3] Department of Public Health Sciences, University of Virginia
ah3wj@virginia.edu, mat3kk@virginia.edu, mb6xn@virginia.edu, achla@virginia.edu, vsakumar@virginia.edu

Abstract

Vaccination is the primary intervention for controlling the spread of infectious diseases. A certain level of vaccination rate (referred to as "herd immunity") is needed for this intervention to be effective. However, there are concerns that herd immunity might not be achieved due to an increasing level of hesitancy and opposition to vaccines. One of the primary reasons for this is the cost of non-conformance with one's peers. We use the framework of network coordination games to study the persistence of anti-vaccine sentiment in a population. We extend it to incorporate the opposing forces of the pressure of conforming to peers, herd-immunity and vaccination benefits. We study the structure of the equilibria in such games, and the characteristics of unvaccinated nodes. We also study Stackelberg strategies to reduce the number of nodes with anti-vaccine sentiment. Finally, we evaluate our results on different kinds of real world social networks.

Introduction

Vaccination is one of the safest methods to control highly contagious childhood diseases, such as measles and smallpox; it is also expected to be one of the primary interventions in controlling the COVID-19 pandemic. The anti-vaccine movement was already becoming an issue for measles in many parts of the US, e.g., in California (Lieu et al. 2015) and Minnesota (Cadena et al. 2019), leading to a concern that the immunization rate may fall below the threshold for herd immunity (Fine 1993; Fine, Eames, and Heymann 2011). This problem has been exacerbated in the COVID-19 pandemic, with some surveys reporting that over 40% of the US population might not take the vaccine, when it becomes available (Cornwall 2020). This poses a significant challenge in the efforts to eliminate the spread of COVID-19.

There are a number of reasons for the fall in immunization rates, but the chief ones are concerns about their possible side-effects, parents' own religious and philosophical beliefs, and misperceptions about the risks (SP 2004; Atwell et al. 2013). It has been observed that peer effects have a significant role in the spread of anti-vaccine sentiment— individuals with such sentiment are in communities with similar sentiment (Poland and Jacobson 2001; Velásquez et al. 2020; Johnson et al. 2020). There is a certain dis-utility an individual gets by not conforming to their social contacts; on the other hand, the individual gets a utility by conforming to its social contacts. This phenomenon can be viewed as a coordination games, which has been very well studied (Ramazi, Riehl, and Cao 2016; Bramoullé and Kranton 2015; Apt, Simon, and Wojtczak 2015; Vanelli et al. 2019; Apt, Simon, and Wojtczak 2019; Jackson and Zenou 2015; Adam, Dahleh, and Ozdaglar 2012). In its basic form, the utility of a node in a coordination game is a function of the number of neighbors having the same state as the node.

In this paper, we extend the framework of coordination games and incorporate vaccination decisions— this requires considering two important components, namely, the benefit a node derives from vaccination, and the benefit of *herd immunity* that all individuals obtain, if a large enough fraction of the population is vaccinated. Studying epidemic decisions from a game-theoretical perspective is not new, e.g. (Bhattacharyya and Ferrari 2017; Shim et al. 2012; Bauch and Earn 2004; Bauch and Bhattacharyya 2012; Aspnes, Chang, and Yampolskiy 2006; V.S. Anil Kumar et al. 2010). However, prior work has either ignored the heterogeneity of human contacts (Bhattacharyya and Ferrari 2017; Shim et al. 2012; Bauch and Earn 2004; Bauch and Bhattacharyya 2012), or the peer effects in vaccine decisions (Aspnes, Chang, and Yampolskiy 2006; V.S. Anil Kumar et al. 2010). In this paper, we study the role of peer effects on the persistence of anti-vaccine clusters in social contact networks. Our contributions are summarized below.

- **Game theoretic approach.** We use a non-cooperative game theoretic approach for studying persistence of anti-vaccine sentiment. We introduce the VACCSENGAME, which extends the framework of coordination games, and incorporates the role of conformity on vaccine sentiment, along with the benefit of vaccination and herd immunity.
- **Structure and complexity of Nash equilibria (NE).** We use NE as the solution concept in such games, and characterize their structure. We show that NE are closely related to the notion of *strong communities* (Flake et al. 2002). We also show a connection between NE and the dynamics of bootstrap percolation (Ackerman, Ben-Zwi, and Wolfovitz 2010; Feige and Kogan 2019), and use it to find the "worst NE", i.e, the one with the largest number of anti-vaccine

nodes. We show that the social optimum (a strategy that maximizes the total utility) can be computed optimally in polynomial time, and derive tight bounds on the Price of Anarchy (the maximum ratio of the total utility of the social optimum and any NE); these terms are defined in the Preliminaries.

- **Empirical analysis.** We study the properties of NE in a diverse class of real-world and social networks and random graphs. We find that there is a threshold value $\theta_{critical}$ for the ratio C/α (where C and α are parameters associated with the benefit from vaccination, and conformity with neighbors, respectively, as defined in the Preliminaries, such that the number of anti-vaccine nodes in the worst NE shows a dramatic change beyond this threshold.
- **Reducing the number of anti-vaccine nodes in the worst NE.** Motivated by our characterization of the worst NE, we explore strategies to reduce the number of anti-vaccine nodes in the worst NE by forcing the decisions for a small subset of nodes. We show that this problem is NP-complete, and find that a heuristic based on high degree is pretty effective.

Implications. The parameter C/α is intuitively reasonable, since C is the parameter capturing the benefit of vaccination, whereas α is the parameter capturing the benefit of conforming to peers. The vaccination benefit C could potentially be increased by suitable incentives, or additional information. The threshold behavior of the number of anti-vaccine nodes in terms of C/α can have an important implication for public health efforts to reduce the number of anti-vaccine nodes: the vaccination benefit has to be increased past a network dependent threshold to give a significant reduction in the number of anti-vaccine nodes.

Preliminaries and Formulation

Network science preliminaries. Let the undirected unweighted graph $G = (V, E)$ denote a network, where V is the set of nodes that represents people in the network and E is the set of edges where each edge represents the connection between two people. Let $n = |V|$ denote the number of nodes in the network. We use $N(i)$ to denote the set of neighbors of node $i \in V$, and $d(i) = |N(i)|$ to denote its degree. For a subset S, we use $d(i, S) = |N(i) \cap S|$ to denote the number of neighbors of i in set S. The graph G represents an "information/influence" network, in which an edge (i, j) means i and j influence each other; G is *not* the network on which the infection spreads in our model.

Bootstrap (threshold) diffusion process (Ackerman, Ben-Zwi, and Wolfovitz 2010; Feige and Kogan 2019). Consider a graph G, and parameter θ_i for each node i. The diffusion process is defined in the following manner:

- Let $S_0 \subseteq V$ be an initial set of nodes
- At timestep $t \geq 1$, each node i which is either in S_{t-1}, or has at least θ_i neighbors in S_{t-1} is added to S_t
- The process stops at time T, when no nodes can be added to S_T, i.e., for $t \geq T : S_{t+1} = S_t$.

Generalizing the notion of *strong communities* (Flake et al. 2002), for vectors $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n_{\geq 0}$, we say $S \subset V$ is a $(\mathbf{p}, \mathbf{q})$-*strong community* (also referred to as $(\mathbf{p}, \mathbf{q})$-strong) if it satisfies the following property: (1) for each node $i \in S$, $|N(i) \cap S| \geq |N(i) - S| + p_i$, and (2) for each $j \notin S$, we have $|N(j) \cap S| \leq |N(j) - S| + q_j$; in the uniform setting, if $p_i = p$ for all i, and $q_j = q$ for all j, we refer to this as a (p, q)-strong community. Intuitively, it means that every node in a strong community has at least as many nodes inside the strong community as outside. We say that S is a *maximal* (p, q)-strong community if it is (p, q)-strong, and there is no $S' \subsetneq S$ that is (p, q)-strong. Finally, we define the *k-core* of G as a maximal subset $S \subset V$ such that $d(i, S) \geq k$ for all $i \in S$.

The VACCSENGAME game. Let $x_i \in \{0, 1\}$ denote the sentiment of node $i \in V$, where 0 and 1 indicate pro- and anti-vaccine sentiments, respectively. For convenience, we refer to a node i as a pro-vaccine or anti-vaccine node, depending on whether $x_i = 0$ or $x_i = 1$, respectively. We use $\mathbf{x}$ to denote the strategy vector. For a subset $S \subset V$, we use $\mathbf{x} = \mathbf{1}_S$ to denote the indicator vector for S, with $x_i = 1$ for $i \in S$, and $x_i = 0$ for $i \notin S$. We use $\texttt{flip}^{(i)}(\mathbf{x})$ to denote the vector obtained by switching only node i's strategy, i.e., $\texttt{flip}^{(i)}(\mathbf{x})_j = x_j$ for $j \neq i$, and $\texttt{flip}^{(i)}(\mathbf{x})_i = 1 - x_i$. Let $N(i, \mathbf{x}, \sigma)$ be the i^{th} node's neighbors who are in state $\sigma \in \{0, 1\}$.

Our VACCSENGAME formulation involves the following components:

- Utility of vaccination, C_i for node i: this parameter captures the utility that a node i gets if it is vaccinated. C_i incorporates both the cost of the vaccine (e.g., economic cost, and both real and rumored side effects to health), as well the (perceived) health benefits; therefore, it can also be negative for some individuals. Since $x_i = 0$ indicates pro-vaccine sentiment, the utility from vaccination for node i can be expressed as $C_i(1 - x_i)$.
- Herd immunity threshold γ, and benefit δ: herd immunity is a standard notion from mathematical epidemiology– if more than a certain fraction (γ) of nodes in the entire network are vaccinated, the disease will die out on its own, e.g., for measles, $\gamma = 0.96$ has been suggested (Bowes 2016). In practice, herd immunity is indirectly considered through the perceived risk of infection. Apart from the individuals ineligible for receiving vaccination, the cost of vaccinations to people who do not have health insurance can be high enough for them to rely on the herd immunity benefit. Such a quantity has actually been determined from analysis of differential equation models, but is given as a general recommendation by public health agencies. We assume that each node gets a benefit δ if the herd immunity threshold is achieved. In terms of our notation of the strategy vector, herd immunity is achieved if $\sum_j x_j < (1 - \gamma)n$, and in this case, each unvaccinated node j gets utility $\delta \cdot x_i \cdot \mathbb{1}_{\sum_j x_j < (1-\gamma)n}$, where $\mathbb{1}_Z$ is an indicator variable which is 1 if the condition Z holds.
- Peer effect parameters $\bar{\alpha}, \bar{\beta}$: we use $\bar{\alpha}$ to indicate the ben-

efit an individual u gets by being "similar" to a neighbor v (i.e., by having the same vaccine sentiment as v). On the other hand, $\bar{\beta}$ is the cost of having a different sentiment from one's neighbor. Without loss of generality (by scaling all the parameters by $1/\bar{\beta}$), we will assume $\bar{\beta} = 1$. Therefore, the utility node i from conforming to its peers is $\bar{\alpha}|N(i, \mathbf{x}, x_i)| - |N(i, \mathbf{x}, 1 - x_i)| = \bar{\alpha}|N(i, \mathbf{x}, x_i)| - (d(i) - |N(i, \mathbf{x}, x_i)|) = (\bar{\alpha} + 1)|N(i, \mathbf{x}, x_i)| - d(i)$. Let $\alpha = \bar{\alpha} + 1$. Since the $-d(i)$ term exists in the utility from conformance for node i, irrespective of its strategy, we drop it, and consider only the term $\alpha|N(i, \mathbf{x}, x_i)|$.

Combining all the above terms, we define the utility for node i as

$$\texttt{util}(i, \mathbf{x}) = \alpha|N(i, \mathbf{x})| + C_i(1 - x_i) + \delta \cdot x_i \mathbb{1}_{\sum_j x_j < (1-\gamma)n}$$

We use $\texttt{util}(\mathbf{x}) = \sum_i \texttt{util}(i, \mathbf{x})$ to denote the total utility associated with $\mathbf{x}$. We refer to a strategy vector $\mathbf{x}^* = \text{argmax}_{\mathbf{x}} \texttt{util}(\mathbf{x})$, which maximizes the total utility, as the *social optimum*.

For notational simplicity, we assume a uniform vaccination utility $C_i = C$ for all i, in our experiments, though the results extend to the general case.

Connection to Coordination games. The standard literature on coordination games (Ramazi, Riehl, and Cao 2016; Bramoullé and Kranton 2015; Apt, Simon, and Wojtczak 2015; Vanelli et al. 2019; Apt, Simon, and Wojtczak 2019; Jackson and Zenou 2015) only consists of the first term, namely $\alpha|N(i, \mathbf{x}, x_i)|$. The second term, namely $C_i(1 - x_i)$ could be easily incorporated into the coordination game setting, e.g., (Vanelli et al. 2019), but is not often considered. The third term, which captures herd immunity has not been considered before. Table 1 summarises the notations used throughout the paper.

Nash equilibrium (NE). We say that $\mathbf{x}$ is a NE if no node i is able to improve its utility by switching its state. In other words, for every i, we have $\texttt{util}(i, \mathbf{x}) \geq \texttt{util}(i, \texttt{flip}^{(i)}(\mathbf{x}))$. We refer to a NE $\mathbf{x}$ which has the maximum number of anti-vaccine nodes as *worst NE*. We define the maximum ratio of the utility of a social optimum to that of any NE, $\frac{\texttt{util}(\mathbf{x}^*)}{\min_{\mathbf{x} NE} \texttt{util}(\mathbf{x})}$ as the *Price of Anarchy* (PoA) (Koutsoupias and Papadimitriou 1999). Note that the definition of (Koutsoupias and Papadimitriou 1999) is the inverse, since they consider cost of a strategy, and the goal is to minimize the cost; since we consider utility maximization, this ratio is more reasonable.

Structural Properties of Nash Equilibria

We observe a correspondence between NE and maximal strong communities.

Lemma 1. *For every $(\mathbf{p}, \mathbf{q})$-strong community S, with $p_i = \frac{C_i}{\alpha}$ and $q_j = \frac{C_j - \delta}{\alpha}$, the state vector $\mathbf{x} = \mathbf{1}_{\bar{S}}$ is a NE. Conversely, if $\mathbf{x}$ is a NE, the set $S = \{i : x_i = 0\}$ is a $(\mathbf{p}, \mathbf{q})$-strong community.*

Symbol	Description
V	Set of vertices
E	Set of edges
n	$\lvert V\rvert$, Number of nodes
$N(i)$	Set of neighbors of node i
$d(i)$	$\lvert N(i)\rvert$, Degree of node i
$d(i, S)$	$\lvert N(i) \cap S\rvert$, Number of neighbors of i in set of nodes S
$x_i(t)$	State of node i at time t, either 0 (pro-vaccine) or 1 (anti-vaccine)
$\mathbf{x}(t)$	Strategy vector at time t
$N(i, \mathbf{x}, \sigma)$	Set of neighbors of node i in $\mathbf{x}$ with state σ
γ	Herd immunity parameter; if $\sum_i x_i < (1 - \gamma)n$ then herd immunity exists
δ	Herd immunity benefit
α	Parameter capturing utility from conformance to neighbors
C_i	Vaccination utility parameter for node i
$\mathbf{1}_S$	Vector $\mathbf{x}$ with $x_i = 1\ \forall i \in S$ and $x_i = 0\ \forall i \notin S$
$\texttt{flip}^{(i)}(\mathbf{x})$	Vector obtained by switching only node i's strategy in $\mathbf{x}$, i.e., $\texttt{flip}^{(i)}(\mathbf{x})_j = x_j$ for $j \neq i$, and $\texttt{flip}^{(i)}(\mathbf{x})_i = 1 - x_i$

Table 1: Summary of notations

Proof. Let S be a $(\mathbf{p}, \mathbf{q})$-strong community S, and $\mathbf{x} = \mathbf{1}_{\bar{S}}$. We observe that no node $i \in V$ improves its utility by switching its strategy, with respect to $\mathbf{x}$. Consider any node $i \in S$. We have $\texttt{util}(i, \mathbf{x}) \geq \alpha|N(i) \cap S|$; note that if the herd immunity condition is met, $\texttt{util}(i, \mathbf{x})$ has an additional δ term, but the above inequality is sufficient for our purpose. Also, $\texttt{util}(i, \texttt{flip}^{(i)}(\mathbf{x})) = \alpha|N(i) - S| + C_i$. By the definition of S, we have $|N(i) \cap S| \geq |N(i) - S| + C_i/\alpha$, which implies $\texttt{util}(i, \mathbf{x}) \geq \texttt{util}(i, \texttt{flip}^{(i)}(\mathbf{x}))$. Next, consider a node $j \notin S$. We have $\texttt{util}(j, \mathbf{x}) = \alpha|N(j) - S| + C_j$ and $\texttt{util}(j, \texttt{flip}^{(j)}(\mathbf{x})) \leq \alpha|N(i) \cap S| + \delta$. By the definition of S, we have $|N(j) \cap S| \leq |N(j) - S| + (C_j - \delta)/\alpha$. Rearranging this we have $\texttt{util}(j, \texttt{flip}^{(j)}(\mathbf{x})) \leq \texttt{util}(j, \mathbf{x})$.

The converse follows by a similar argument. □

Connection to bootstrap percolation. For the pure coordination game, the best response strategy is closely related to the threshold model of diffusion or bootstrap percolation (Adam, Dahleh, and Ozdaglar 2012; Barrett et al. 2006). We show below that a similar connection exists for VACCSENGAME, with thresholds which depend on the vaccination and herd immunity parameters.

Lemma 2. *Let $S \subseteq V$ be any non empty subset of nodes. Let $U = S_T$ be the final set obtained by running the threshold process with $\theta_v = \frac{d(v)}{2} - \frac{C_v - \delta}{2\alpha}$, starting from $S_0 = S$, and let $\mathbf{x} = \mathbf{1}_{V-U}$. If $|U| > \gamma n$, or if $\delta = 0$, no node $v \in V - S$ has incentive to switch its strategy.*

Proof. Let S_t $(t = 0, \ldots, T)$ be the sequence of sets in the diffusion process, with $S_0 = S$ and $S_T = U$.

First, we consider the case $|U| > \gamma n$. Consider any node $v \in S_t - S_{t-1}$, $t \geq 1$. By definition of the diffusion process, we have $d(v, S_{t-1}) \geq \theta_v = \frac{d(v)}{2} - \frac{C_v - \delta}{2\alpha}$. Rearranging the terms, we have $\alpha d(v, S_{t-1}) + C_v \geq \alpha(d(v) - d(v, S_{t-1})) + \delta = \alpha d(v, \overline{S_{t-1}}) + \delta$. Further, we have $d(v, U) \geq d(v, S_{t-1})$ and $d(v, \overline{S_{t-1}}) \geq d(v, \overline{U})$, as $S_{t-1} \subseteq U$. This implies $\texttt{util}(v, \mathbf{x}) = \alpha d(v, U) + C_v \geq \alpha d(v, \overline{U}) + \delta = \texttt{util}(v, \texttt{flip}^{(v)}(\mathbf{x}))$, so that node v has no utility to flip its strategy.

Next, consider any node $v \notin U$. By definition of the diffusion process, we have $d(v, U) < \theta_v = \frac{d(v)}{2} - \frac{C_v - \delta}{2\alpha}$. Rearranging the terms, we have $\alpha d(v, U) + C_v < \alpha(d(v) - d(v, U)) + \delta = \alpha d(v, \overline{U}) + \delta$, which implies $\texttt{util}(v, \texttt{flip}^{(v)}(\mathbf{x})) < \texttt{util}(i, \mathbf{x})$. This means node v does not have incentive to switch its strategy in this case either.

The argument in the case $\delta = 0$ is similar. □

Theorem 3. *Suppose $\delta = 0$. Let $S_0 = \{v : d(v) < \frac{C_v}{\alpha}\}$. Let U be the final set obtained by running the threshold process with $\theta_v = \frac{d(v)}{2} - \frac{C_v}{2\alpha}$, starting from S_0. Then, the strategy vector $\mathbf{x} = \mathbf{1}_{V-U}$ is a pure NE. Further, this is the NE with the minimum number of pro-vacc nodes.*

Proof. From Lemma 2, it follows that no node $v \in V - S_0$ has incentive to switch its strategy. Consider a node $v \in S_0$. We have $d(v) < \frac{C_v}{\alpha}$, which implies $d(v) < 2d(v, U) + \frac{C_v}{\alpha}$. Rearranging, we have $d(v) - d(v, U) < d(v, U) + \frac{C_v}{\alpha}$, so that $\texttt{util}(v, \texttt{flip}^{(v)}(\mathbf{x})) = \alpha d(v, \overline{U}) < \texttt{util}(v, \mathbf{x}) = \alpha d(v, U) + C_v$, so that node v has no incentive to switch its strategy. Therefore, $\mathbf{x}$ is a NE.

The above argument also implies that in any NE $\mathbf{x}$, the nodes in S_0 will always have state 0. An inductive argument, as in the proof of Lemma 2, also implies that each node $v \in S_t$ has incentive to switch to state 0. This implies that in any NE, the nodes in U are always pro-vaccine, and the Lemma follows. □

We refer to the NE in Theorem 3 as the worst NE.

Finding Nash Equilibria

It is known that a best response strategy converges to a NE in pure coordination games (Bramoullé and Kranton 2015; Apt, Simon, and Wojtczak 2015; Vanelli et al. 2019; Apt, Simon, and Wojtczak 2019; Ramazi, Riehl, and Cao 2016); some of these analyses can also be extended to prove that when $\delta = 0$, a best response type of strategy converges to a NE. We show that a best response strategy, SEQBR, in a specific order converges much faster to a NE when $\delta = 0$, and can be implemented in $O(|V| + |E|)$ time. In our experiments, we find a parallel best response, PARBR, converges faster. We give details of these methods, and their analyses in the Appendix. When $\delta > 0$, finding a NE in which herd immunity is achieved is much harder, as we discuss below.

Lemma 4. *Let $K < (1 - \gamma)n$. Determining whether or not there exists a NE $\mathbf{x}$ with $0 < \sum_i x_i \leq K$ is NP-complete.*

The Social Optimum and the Price of Anarchy

Recall the notions of social optimum and price of anarchy, as defined in the Preliminaries. We first bound the PoA.

Theorem 5. *For any instance with $\delta = 0$, and $C_i \geq 0$ for all i, the PoA is at most $2 + \frac{\sum_i C_i}{\alpha|E|}$.*

Proof. First, observe that $\sum_i \texttt{util}(i, \mathbf{x}^*) = \sum_i \alpha|N(i, \mathbf{x}^*)| + \sum_i C_i(1 - x_i^*) \leq 2\alpha|E| + \sum_i C_i$.

Next, consider any NE $\mathbf{x}$. Let $U = \{i : x_i = 0\}$. By definition of a NE, for $i \in U$, we have $\alpha|N(i) \cap U| + C_i \geq \alpha|N(i) \cap \overline{U}| = \alpha(d(i) - |N(i) \cap U|)$. Rearranging, we have $\alpha|N(i, U)| + C_i/2 \geq \alpha d(i)/2$. Similarly, for $i \in \overline{U}$, we have $\alpha|N(i) \cap \overline{U}| \geq \alpha|N(i) \cap U| + C_i \geq \alpha|N(i) \cap U|$. Rearranging, we have $\alpha|N(i) \cap \overline{U}| \geq \alpha d(i)/2$. This implies $\sum_i \texttt{util}(i, \mathbf{x}) = \sum_{i \in U} \alpha|N(i) \cap U| + C_i + \sum_{i \in \overline{U}} \alpha|N(i) \cap \overline{U}| \geq \sum_{i \in U} \alpha|N(i) \cap U| + C_i/2 + \sum_{i \in \overline{U}} \alpha|N(i) \cap \overline{U}| \geq \alpha \sum_i d(i)/2 = \alpha|E|$.

Putting these together, the PoA is bounded by $\frac{2\alpha|E| + \sum_i C_i}{\alpha|E|} \leq 2 + \frac{\sum_i C_i}{\alpha|E|}$. □

Next, we show that the social optimum can be computed in polynomial time using an approach based on linear programming and rounding. Our algorithm involves the following steps.

1. Solve the following linear program (LP)

$$\max \quad \sum_i C_i(1 - z_i) + \alpha \sum_e (1 - y_e) \tag{1}$$
$$y_e \geq z_i - z_j \text{ for all } e = (i, j) \in E \tag{2}$$
$$y_e \geq z_j - z_i \text{ for all } e = (i, j) \in E \tag{3}$$
$$z_i, y_e \in [0, 1] \text{ for all } i \in V, e \in E \tag{4}$$

2. Let y, z be an optimal factional solution to the above linear program
3. Let $S = \{z_i : i \in V\}$. For each $r \in S$, define the strategy vector $\mathbf{x}(r)$ with $x_i(r) = 1$ if $z_i \geq r$. Return the vector $\mathbf{x}(r^*) = \text{argmax}_{r \in S} \sum_i \texttt{util}(i, \mathbf{x}(r))$, which maximizes the total utility.

Theorem 6. *If $\delta = 0$, the strategy $\mathbf{x}(r^*)$ computed by the above algorithm is a social optimum.*

Proof. (Sketch) Let $\mathbf{x}^*$ denote a social optimum. First observe that in the linear program (LP), we have $y_e = |z_i - z_j|$ for all $e = (i, j)$. This follows because the objective involves maximizing $\sum_e (1 - y_e)$. If there exists an edge e with $y_e > |z_i - z_j|$, we can reduce y_e and increase the objective value, while keeping the z_i's fixed.

Next, we observe that the integral version of the above linear program, i.e., with the constraints (4) replaced with $z_i, y_e \in \{0, 1\}$, has the same total utility as the social optimum $\mathbf{x}^*$. Consider a solution $z_i = x_i^*$, and for $e = (i, j)$, $y_e = 1$ if $x_i^* \neq x_j^*$. Then, observe that $y_e = |z_i - z_j|$, and so y, z is a feasible solution. Let $S^* = \{i : x_i^* = 0\}$. This implies for any node $i \in S^*$, $|N(i) \cap S^*| = \sum_{e=(i,j), j \in S^*} (1 - y_e)$. Similarly, for node $i \in \overline{S^*}$, $|N(i) \cap$

$\overline{S^*}| = \sum_{e=(i,j), j\in\overline{S^*}}(1-y_e)$. Therefore, the objective value of y, z equals $\texttt{util}(\mathbf{x}^*)$.

We first consider a different rounding than the one in Step 3 of the algorithm: pick $r \in [0,1]$ uniformly at random, and let $\mathbf{X}$ denote the strategy vector with $X_i = 1$ if $z_i \geq r$. Let $Y_e = |X_i - X_j|$. Then, $E[X_i] = \Pr[X_i = 1] = \Pr[r \leq z_i] = z_i$. Further, $Y_e = 1$ if $r \in (\min(X_i, X_j), \max(X_i, X_j))$. This implies $E[Y_e] = |z_i - z_j| = y_e$. By linearity of expectation, we have $E[\texttt{util}(\mathbf{X})] = \alpha\sum_e(1-Y_e) + \sum_i C_i(1-X_i) = \texttt{util}(\mathbf{x}^*)$. This implies there exists a value of $r = r^*$ such that the strategy $\mathbf{x}(r^*)$ defined in Step 3 of the algorithm has $\texttt{util}(\mathbf{x}(r^*)) = E[\texttt{util}(\mathbf{X})]$, and so the strategy $\mathbf{x}(r^*)$ is a social optimum. □

Interventions to Reduce the Number of Anti-vaccine Nodes in the Worst NE

In general, the worst NE can have a large fraction of anti-vaccine nodes. For instance, consider an instance with $d(v) > C_v/\alpha$ for all v. Then, by Theorem 3, the worst NE has n anti-vaccine nodes. This motivates the following question: can we incentivize a set S of at most k nodes to become pro-vaccine, so that if the remaining nodes in $V - S$ make decisions maximizing their individual utility, the number of anti-vaccine nodes is minimized; here k denotes the budget available to the social planner. This approach falls into the framework of *Stackelberg* strategies (Roughgarden 2004), in which a social planner is able to force the strategies for a subset of players, and the rest decide in a decentralized manner. We refer to the set S as a Stackelberg solution. We say a NE $\mathbf{x}^S$ is consistent with a Stackelberg solution S if $x_v^S = 0$ for all $v \in S$, and for all $i \notin S$, we have $\texttt{util}(i, \mathbf{x}^S) \geq \texttt{util}(i, \texttt{flip}^{(i)}(\mathbf{x}^S))$, i.e., no node in $\overline{S}$ has incentive to switch its decision.

Lemma 7. *Given a budget k and a target value N, finding a Stackelberg solution S with $|S| \leq k$, such that the number of anti-vaccine nodes in the worst NE $\mathbf{x}^S$ is at most N, is NP-complete.*

Proof. (Sketch) We do a reduction from the Target Set Selection in threshold dynamical systems (Ackerman, Ben-Zwi, and Wolfovitz 2010). Given a network G and a threshold θ_v for each node $v \in V$, the objective is to choose an initial set S_0, so that all the nodes are influenced eventually. Determining whether there exists a solution S_0 with $|S_0| \leq k$, such that all nodes are influenced is NP-complete. Using Theorem 6, it follows that finding a Stackelberg solution S such that the number of anti-vaccine nodes is $N = 0$ is NP-complete. □

Finding good Stackelberg solutions. From Lemma 7, it follows that this problem is computationally very hard. For the special case of $C = 0$, the techniques of (Feige and Kogan 2019) can be adapted to give an $O(\max_v d(v))$-approximation. Since this is not a very practical algorithm, we use a degree based strategy as a heuristic to find a Stackelberg solution: given a budget k, we pick the top k nodes in terms of degree, and add to the set S_0 in Theorem 6, and find the worst NE.

Network	$n = \|V\|$	$\|E\|$	d_{avg}	cc_{avg}	$\theta_{critical}$
Synthetic					
Erdos-Renyi	1000	99851	199.702	0.1998	169.0
Social					
Facebook combined	4039	88234	43.691	0.6055	93.0
Twitter mentions	9527	82709	17.363	0.0505	136.0
Communication					
email-Eu-core	1005	16706	33.2458	0.3994	17.0
Gnutella p2p	10876	39994	7.3545	0.0062	3.0

Table 2: Summary of datasets– number of nodes (n), number of edges ($|E|$), average degree (d_{avg}) and average clustering coefficient (cc_{avg}). $\theta_{critical}$ is the threshold values beyond which the number of anti-vaccine nodes in the worst NE suddenly decline to zero . Refer to Section for the experiment details. We mention one of the Erdős-Rényi used here; we also consider other values of p. The social and communication networks were collected from the Stanford Network Analysis Project (Leskovec and Krevl 2014).

Experimental Results

We study the following questions to complement our theoretical results and understand the structure of VACCSENGAME in different kinds of networks.

- **Minimizing the number of anti-vaccine nodes in the worst NE:** how does the number of anti-vaccine nodes in the worst NE vary with the parameter C/α (which affects the initial set of pro-vaccine nodes S_0 in Theorem 3 to compute worst NE) in different graphs, and at what point does it drop significantly?
- **Characteristics of nodes in worst and random NE.** Do properties such as degree and clustering coefficient help characterize the anti-vaccine nodes?
- **Effectiveness of strategies to limit worst NE.** Can a small set of nodes be influenced, so that the number of anti-vaccine nodes in the worst NE drops?

We perform all the experiments using Python 3.7.5 on a Windows 10 Pro machine with 16 GB of physical memory. Networkx was used for graph manipulation, Pandas and Numpy libraries were used for data analysis and Matplotlib for visualization.

Datasets

We consider a variety of synthetic graphs and real-world networks as summarized in Table 2. We also analyze our own Twitter network from data collected in June 2020 (using the public API), allowing for an experiment representative of online discussion in response to COVID-19. This network differs from the rest in that it has directed edges along with node attributes that are an aggregation of a user's sentiment from text. Some results are omitted due to space, and will be presented in the full version of the paper.

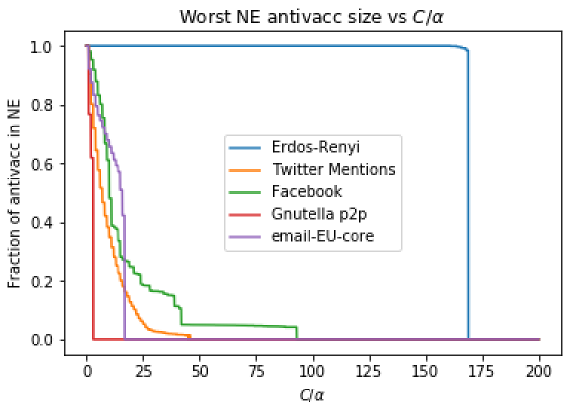

Figure 1: Fraction of nodes which end up as anti-vaccine in the worst NE according to Theorem 3. We observe a threshold effect where the anti-vaccine nodes disappear as C/α increases.

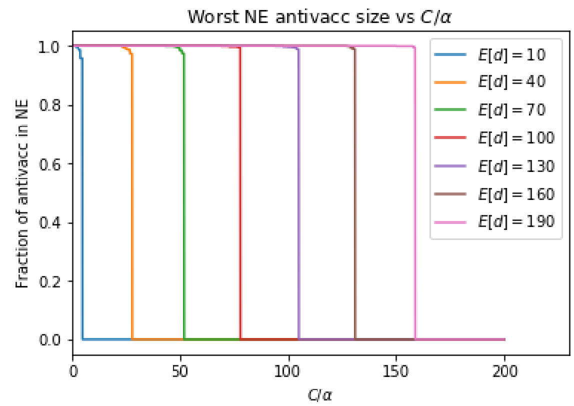

Figure 2: Fraction of nodes which end up as anti-vaccine in the worst NE according to Theorem 3 for the Erdős-Rényi model $G(n, p)$ for different values of p. As we increase the expected degree np in the graph, the threshold value $\theta_{critical}$ increases.

Number of Anti-vaccine Nodes in the Worst NE

We compute the worst NE for the datasets summarized in Table 2 using the approach outlined in Theorem 3. We use a uniform value for C here. Figure 1 shows the number of anti-vaccine nodes in the worst NE (y-axis) as a function of the parameter C/α which plays a role in the choice of the set S_0 in Theorem 3, and intuitively captures the ratio of the benefit of vaccination to that of conformance with peers. As expected, the number of anti-vaccine nodes reduces when C/α increases. Quite surprisingly, we observe that the number of anti-vaccine nodes exhibits a "threshold phenomenon" with respect to C/α, i.e., a small increase in C/α leads to a significant drop in the number of anti-vaccine nodes. We refer to the value of C/α at which number of anti-vaccine drops to zero as $\theta_{critical}$. The $\theta_{critical}$ values are highly network dependent. In particular, the $G(n, p)$ model of Erdős-Rényi has a very sharp threshold. In contrast, the Facebook and Twitter networks have a sharp initial drop, followed by some spread.

In Figure 2, we examine the threshold phenomenon in the $G(n, p)$ model for other p values; the expected degree np is shown for each plot. $\theta_{critical}$ seems to be a constant fraction of the expected degree (within a factor of 2); understanding the exact dependence is an interesting open problem.

The vaccination benefit C could be increased by suitable incentives, or additional information. The threshold effect relative to the parameter C/α suggests that raising the vaccination benefit past the threshold can have a significant public health benefit. In contrast, increasing C/α below this threshold does not have a significant benefit for reducing the number of anti-vaccine nodes.

Characteristics of Nodes in the Worst NE

We consider the worst NE associated with the C/α values just before the threshold, and examine the characteristics of nodes in these NE. Figure 3 shows the degree and clustering coefficient distributions of the anti-vacc nodes in the NE for the Twitter mentions network, just before the critical threshold. We observe that these nodes have somewhat low degrees, and generally low clustering coefficients. Similar results for other networks are presented in the Appendix.

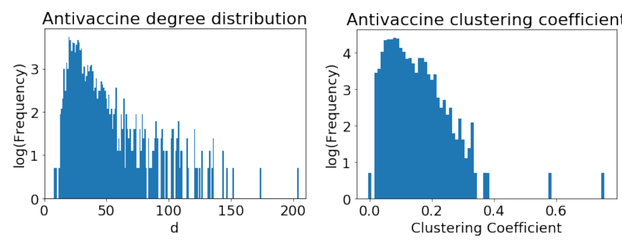

Figure 3: (Left) Degree distribution, and (Right) Clustering coefficient distribution of provaccine nodes in the worst NE in the Twitter network for $C/\alpha = 10$ (before $\theta_{critical}$).

Characteristics of Random NE

VACCSENGAME has multiple NE, in general, and here we examine the characteristics of the anti-vaccine nodes, specifically how they differ from those in the worst NE. Figure 4 shows the degree distribution and the clustering coefficient distribution of the combined Twitter network's anti-vaccine nodes in a NE computed using a best response strategy. In the initial strategy vector (for the best response), the anti-vaccine nodes are chosen randomly. We set $\alpha = \delta = 1$ and $\gamma = 0.9$. We observe that in contrast to Figure 3, nodes have much lower degree as well as clustering coefficient.

Reducing the Number of Anti-vaccine Nodes in Worst NE

Here, we study the effectiveness of Stackelberg strategies for reducing the number of anti-vaccine nodes. Motivated by Lemma 2, we use strategies to select the initial set S_0, so that the number of anti-vaccine nodes in the resulting

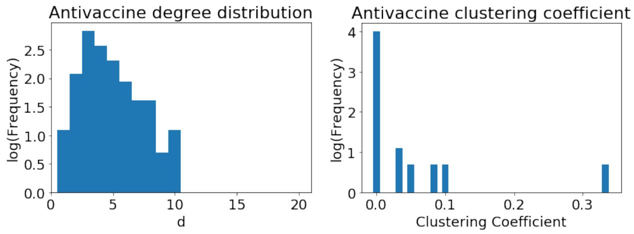

Figure 4: (Left) Degree distribution and (Right) clustering coefficient distribution of the combined Twitter network's anti-vaccine nodes in a NE. Note that the box plot in the left represents the distribution of the number of nodes for the corresponding interval of degrees.

worst NE is minimized. Let $\mathbf{D}_k$ be the set of k nodes of the highest degree. We choose $S_0 = \{v : d_v < C_v/\alpha\} \cup \mathbf{D}_k$. In Figure 5 we observe significant variation in $\theta_{critical}$ with $\frac{k}{n}$ across networks but for all networks $\theta_{critical}$ goes down very sharply as $\frac{k}{n}$ increases.

Related Work

Our paper is a direct generalization of coordination games, which have been studied extensively, e.g., (Apt, Simon, and Wojtczak 2015; Vanelli et al. 2019; Apt, Simon, and Wojtczak 2019; Ramazi, Riehl, and Cao 2016; Adam, Dahleh, and Ozdaglar 2012); we refer to (Jackson and Zenou 2015; Bramoullé and Kranton 2015) for good surveys on this topic, and on games on networks, more broadly. As mentioned earlier, such games typically only involve the peer effects term from VACCSENGAME, which corresponds to setting C_i and δ to 0. Prior work has primarily focused on the convergence of both synchronous and asynchronous best response strategies (Ramazi, Riehl, and Cao 2016); they show convergence in linear time. (Vanelli et al. 2019) analyze both coordination and anti-coordination games (in which nodes prefer to have a different state than their neighbors), and derive tight bounds on the size of different NE; however, these are restricted to complete networks. (Apt et al. 2014) consider a stronger notion of NE, and present conditions for their existence. These games have also been extended to weighted (Apt, Simon, and Wojtczak 2019) and directed networks (Apt, Simon, and Wojtczak 2015); they show that NE need not exist, in general, and finding them is NP-complete. (Adam, Dahleh, and Ozdaglar 2012) also show a connection between best response and cascades in the bootstrap model.

There has also been a lot of game-theoretic work on vaccination decisions, which have considered the costs of vaccination and sickness, e.g., (Bhattacharyya and Ferrari 2017; Shim et al. 2012; Bauch and Earn 2004; Bauch and Bhattacharyya 2012; Aspnes, Chang, and Yampolskiy 2006; V.S. Anil Kumar et al. 2010). Most of this has used differential equation models, which makes it easier to incorporate the cost of infection (Bhattacharyya and Ferrari 2017; Shim et al. 2012; Bauch and Earn 2004; Bauch and Bhattacharyya 2012). There has been limited work on vaccination games in network models (Aspnes, Chang, and Yampolskiy 2006; V.S. Anil Kumar et al. 2010). However, these works have not

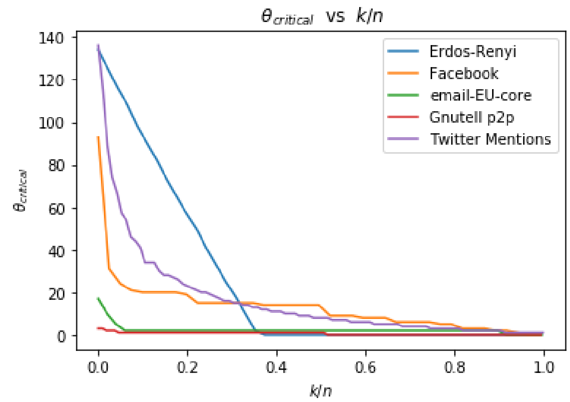

Figure 5: Variation in $\theta_{critical}$ (y-axis) vs $\frac{k}{n}$ (x-axis) for different networks.

taken the important role that peer effects play in such such decisions. Our work is the first to extend coordination games to incorporate the benefits of vaccination and herd immunity.

There is a lot of work on the broader topic of vaccine sentiment, and how it can be changed (Velásquez et al. 2020; Johnson et al. 2020; Cornwall 2020). Its been found that polarization and strong ideologies make people less amenable to changing their sentiment; in particular, providing information on vaccine safety and benefits might have an opposite effect on some people.

Conclusions

In this study, we extend the framework of coordination games to study the spread of anti-vaccine sentiment in a social network. We study the structure of Nash equilibria, including the maximum number of anti-vaccine nodes, and their characteristics. Quite surprisingly, we find that the ratio C/α, which captures the relative benefit of vaccination to conformance to one's peers has a threshold effect on the number of anti-vaccine nodes. The point at which the threshold occurs is network dependent, and understanding is an interesting open question. This can also help in designing incentives to reduce anti-vaccine sentiment. This is a fundamental public health challenge, as vaccinations are the only hope for eliminating highly contagious diseases, including COVID-19.

Our analysis can be extended in multiple directions. The price-demand relationship of vaccines is an important economic aspect of this problem. If the number of people taking the vaccines increases, the cost associated with taking the vaccines will also increase. Incorporating this into VACCSENGAME is an interesting future direction, specially in the context of potential competing vaccinations with different costs and efficacies. Another direction is to consider the hardening of ideologies when information or incentives for vaccination is provided (Velásquez et al. 2020; Johnson et al. 2020; Cornwall 2020). This issue needs to be considered carefully in the Stackelberg approach.

Acknowledgments

This work was partially supported by National Institutes of Health (NIH) Grant 2R01GM109718 and 3R01GM109718, NSF grants IIS-1931628, CCF-1918656, IIS-1955797, IIS-2027848, DTRA subcontract/ARA S-D00189-15-TO-01-UVA, and UVA grant 165924.101.LC00186.30001. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies.

References

Ackerman, E.; Ben-Zwi, O.; and Wolfovitz, G. 2010. Combinatorial model and bounds for target set selection. *Theoretical Computer Science* 411(44): 4017 – 4022. ISSN 0304-3975. doi:https://doi.org/10.1016/j.tcs.2010.08.021. URL http://www.sciencedirect.com/science/article/pii/S0304397510004561.

Adam, E. M.; Dahleh, M. A.; and Ozdaglar, A. 2012. On the behavior of threshold models over finite networks. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, 2672–2677.

Apt, K. R.; Rahn, M.; Schäfer, G.; and Simon, S. 2014. Coordination Games on Graphs (Extended Abstract). In Liu, T.; Qi, Q.; and Ye, Y., eds., *Web and Internet Economics - 10th International Conference, WINE 2014, Beijing, China, December 14-17, 2014. Proceedings*, volume 8877 of *Lecture Notes in Computer Science*, 441–446. Springer.

Apt, K. R.; Simon, S.; and Wojtczak, D. 2015. Coordination Games on Directed Graphs. In Ramanujam, R., ed., *Proceedings Fifteenth Conference on Theoretical Aspects of Rationality and Knowledge, TARK 2015, Carnegie Mellon University, Pittsburgh, USA, June 4-6, 2015*, volume 215 of *EPTCS*, 67–80. doi:10.4204/EPTCS.215.6. URL https://doi.org/10.4204/EPTCS.215.6.

Apt, K. R.; Simon, S.; and Wojtczak, D. 2019. Coordination Games on Weighted Directed Graphs. *CoRR* abs/1910.02693. URL http://arxiv.org/abs/1910.02693.

Aspnes, J.; Chang, K.; and Yampolskiy, A. 2006. Inoculation strategies for victims of viruses and the sum-of-squares partition problem. *J. Comput. Syst. Sci.* .

Atwell, J. E.; Otterloo, J. V.; Zipprich, J.; Winter, K.; Harriman, K.; Salmon, D. A.; Halsey, N. A.; and Omer, S. B. 2013. Nonmedical Vaccine Exemptions and Pertussis in California, 2010. *Pediatrics* .

Barrett, C. L.; Hunt, H. B.; Marathe, M. V.; Ravi, S.; Rosenkrantz, D. J.; and Stearns, R. E. 2006. Complexity of reachability problems for finite discrete dynamical systems. *Journal of Computer and System Sciences* 72(8): 1317 – 1345. ISSN 0022-0000. doi:https://doi.org/10.1016/j.jcss.2006.03.006. URL http://www.sciencedirect.com/science/article/pii/S002200000600047X.

Bauch, C. T.; and Bhattacharyya, S. 2012. Evolutionary game theory and social learning can determine how vaccine scares unfold. *PLoS computational biology* 8(4).

Bauch, C. T.; and Earn, D. J. 2004. Vaccination and the theory of games. *Proceedings of the National Academy of Sciences* 101(36): 13391–13394.

Bhattacharyya, S.; and Ferrari, M. 2017. Age-specific mixing generates transient outbreak risk following critical-level vaccination. *Epidemiology and Infection* 145(1): 12.

Bowes, J. 2016. Measles, misinformation, and risk: personal belief exemptions and the MMR vaccine. *Journal of law and the biosciences* 718–725.

Bramoullé, Y.; and Kranton, R. 2015. Games Played on Networks. Working Papers halshs-01180657, HAL. URL https://ideas.repec.org/p/hal/wpaper/halshs-01180657.html.

Cadena, J.; Falcone, D.; Marathe, A.; and Vullikanti, A. 2019. Discovery of under immunized spatial clusters using network scan statistics. *BMC Medical Informatics and Decision Making* 19(1): 28.

Cornwall, W. 2020. Just 50% of Americans plan to get a COVID-19 vaccine. Here's how to win over the rest. *Science* https://www.sciencemag.org/news/2020/06/just-50-americans-plan-get-covid-19-vaccine-here-s-how-win-over-rest.

Feige, U.; and Kogan, S. 2019. Target Set Selection for Conservative Populations. *ArXiv* abs/1909.03422.

Fine, P.; Eames, K.; and Heymann, D. L. 2011. "Herd immunity": a rough guide. *Clinical infectious diseases* 52(7): 911–916.

Fine, P. E. 1993. Herd immunity: history, theory, practice. *Epidemiologic reviews* 15(2): 265–302.

Flake, G. W.; Lawrence, S.; Giles, C. L.; and Coetzee, F. M. 2002. Self-organization and identification of Web communities. *Computer* 35(3): 66–70. ISSN 1558-0814. doi:10.1109/2.989932.

Jackson, M. O.; and Zenou, Y. 2015. Games on networks. In *Handbook of game theory with economic applications*, volume 4, 95–163. Elsevier.

Johnson, N.; Velasquez, N.; Restrepo, N.; Leahy, R.; Gabriel, N.; Oud, S.; Zheng, M.; Manrique, P.; Wuchty, S.; and Lupu, Y. 2020. The online competition between pro- and anti-vaccination views. *Nature* 582. doi:10.1038/s41586-020-2281-1.

Koutsoupias, E.; and Papadimitriou, C. 1999. Worst-case equilibria. In *STACS '99*, 404–413.

Leskovec, J.; and Krevl, A. 2014. SNAP Datasets: Stanford Large Network Dataset Collection. http://snap.stanford.edu/data.

Lieu, T. A.; Ray, G. T.; Klein, N. P.; Chung, C.; and Kulldorff, M. 2015. Geographic clusters in underimmunization and vaccine refusal. *Pediatrics* 135(2): 280–289.

Poland, G. A.; and Jacobson, R. M. 2001. Understanding those who do not understand: a brief review of the anti-vaccine movement. *Vaccine* 19(17-19): 2440–2445.

Ramazi, P.; Riehl, J.; and Cao, M. 2016. Networks of conforming or nonconforming individuals tend to reach satisfactory decisions. *Proceedings of the National Academy of Sciences* 113(46): 12985–12990.

Roughgarden, T. 2004. Stackelberg Scheduling Strategies. *SIAM J. Comput.* 33(2): 332–350.

Shim, E.; Grefenstette, J.; Albert, S.; Cakouros, B.; and Burke, D. 2012. A game dynamic model for vaccine skeptics and vaccine believers: measles as an example. *J Theor Biol.* 194–203.

SP, C. 2004. Vanishing vaccinations: why are so many Americans opting out of vaccinating their children? *Univ Mich J Law Reform* 353–440.

Vanelli, M.; Arditti, L.; Como, G.; and Fagnani, F. 2019. On games with coordinating and anti-coordinating agents. *CoRR* abs/1912.02000. URL http://arxiv.org/abs/1912.02000.

Velásquez, N.; Leahy, R.; Restrepo, N.; Lupu, Y.; Sear, R.; Gabriel, N.; Jha, O.; Goldberg, B.; and Johnson, N. F. 2020. Hate multiverse spreads malicious COVID-19 content online beyond individual platform control. *arXiv: Physics and Society* .

V.S. Anil Kumar; Rajaraman, R.; Sun, Z.; and Sundaram, R. 2010. Existence Theorems and Approximation Algorithms for Generalized N etwork Security Games. In *Proceedings of IEEE ICDCS*.