Trust in Persuasion for Binary Adversarial Classification

Reema Deori
Centre for Systems and Control
IIT Bombay, Mumbai, India
deori.reema@iitb.ac.in

Ankur A. Kulkarni
Centre for Systems and Control
IIT Bombay, Mumbai, India
kulkarni.ankur@iitb.ac.in

***Abstract*— A sender has access to a dataset which is split into two classes. The sender observes these data points which get generated uniformly at random and sends the receiver a message about its observation. The receiver's goal is to correctly classify the data point from the message sent by the sender. The sender wants to persuade the receiver to choose a class that maximizes the sender's utility which is dependent on how a data point is classified. We study this interaction as a *Stackelberg game*. We show that if the sender wants to get some proportion of data points of a particular label misclassified using a message, then at every *Stackelberg equilibrium strategy*, it should use the same message for equal or more proportion of data points whose true label is that particular label. We study two special classes of utility functions and show that the *obscureness* of the sender, which is the maximum number of data points that can be misclassified at equilibrium, is equal to the total number of data points of the least frequent label in the dataset. In addition to quantifying obscureness, we also quantify the minimum information that needs to be revealed by the sender at Stackelberg equilibrium, which we call the *informativeness* of the sender. We prove these results by characterizing the Stackelberg equilibrium game and the informativeness of the sender via two *linear programs* we introduced in this paper. The linear program characterization is the key contribution of our paper.**

***Index Terms*— Game theory, Bayesian Persuasion, Stackelberg equilibrium, binary adversarial classification, trust, semantic communication**

I. Introduction

Suppose that our sender is a bank that sends multiple emails to its customers. Not all emails sent by the bank are of importance or urgent. A customer would only want to read or open emails that are informative or urgent like reminders for paying insurance before the due date, transaction details, and so on. But in addition to sending urgent emails, the bank also needs to promote its new schemes and other such additional information that are of a non-urgent nature to the customer. The bank is aware of the true nature of the content of the emails it sends out to its customers and uses one or more email addresses for this purpose. A customer is well aware of this and comes up with a filtering or classification policy that will maximize the total number of correctly classified emails sent from the email address used. But the bank being a strategic agent must come up with a way of persuading the customer to classify a non-urgent email as urgent. Imagine a situation where the proportion of non-urgent emails is more than the urgent ones. In such a scenario, if the bank resorts to the use of a single email address for sending all these emails, then the customer will classify all emails as non-urgent which hurts both the customer and the bank. By using distinct email addresses, the bank can persuade the customer to read non-urgent emails. To ensure that the majority of the emails sent from a particular email address are correctly labeled as urgent emails, the proportion of urgent emails sent from that email address can never be less than the proportion of non-urgent emails sent using the same email address. This persuades the customer to misclassify some non-urgent emails as urgent when sent from that particular email address. This kind of strategic signalling by the bank ensures that not all non-urgent emails are filtered out or categorized as non-urgent.

The email filtering example hints that the only way the bank could convince a customer to label a particular email address as non-urgent is by providing him enough evidence to persuade it into believing that it is in the best interest of the customer to label that particular address as non-urgent. Thus, an intuitive conclusion that one can draw from this example is that an agent can be persuaded into picking an action preferred by the sender only if the sender can gain the trust of the agent by presenting the information in such a way that the agent is convinced that the sender's preferred action is one of the best possible action for the agent in that scenario.

Using this example as motivation we study a model which captures this kind of strategic interaction between a sender and a receiver, the sender has access to a dataset and has knowledge of the true label of the data points in the dataset. The sender can persuade the receiver that it has observed some data point by sending a unique signal for every data point or a single signal for multiple data points. The receiver's goal is to maximize the proportion of data points that are correctly classified or labeled. Therefore, after observing the sender's strategy, the receiver then picks a class for every observed signal to maximize the proportion of correctly classified data points. We pose this problem as a *Stackelberg game* with the sender as the leader and characterize the Stackelberg equilibrium strategies. By studying these Stackelberg equilibrium strategies, we conclude how much information needs to be revealed by the sender to maximize its expected utility.

ISBN 978-3-903176-65-2

A. Main results

In a Stackelberg game, once the sender declares its choice of strategy, the receiver will pick a strategy in response which maximizes its expected utility. We call such strategies of the receiver *best response strategies*. The maximum possible expected utility attainable by the sender when the receiver only plays one of its best response strategies is referred to as the *Stackelberg game value* in our paper. A strategy of the sender that gets the sender a payoff equal to Stackelberg game value is called *Stackelberg equilibrium strategies* of the sender. The strategy spaces of the players are probability spaces which makes the strategy spaces infinite in size. For computing the Stackelberg game value, one needs to compute the best-case expected utility for every strategy of the sender which is computationally expensive. Our main contribution is the formulation of a linear program (LP) whose optimal value is exactly equal to the Stackelberg game value. Thus, solving the LP for computing the Stackelberg game value makes it a computationally less expensive task when compared to a brute-force approach. The LP not only gives us the Stackelberg game value but the optimal solutions of the LP also help us identify the Stackelberg equilibrium strategies of the sender. We show that every optimal solution of the LP is equivalent to some Stackelberg equilibrium strategy of the sender and vice-versa. Next, we also study two particular classes of utility functions of the sender and characterize their Stackelberg game values. We concluded that for these special class of utility functions, the maximum number of data points that can be misclassified to maximize the sender's utility called *obscureness* is exactly equal to the number of data points in the dataset of the least frequent label. In addition to this, we also quantified the minimum number of symbols that are correctly classified at equilibrium. We call this quantity *informativeness* and we show that informativeness can also be computed by solving another LP for the given pair of true labeling function and utility function of the sender. We also computed the informativeness for the two special classes of utility functions. We concluded that the informativeness of the sender for these two classes of utility functions is the number of data points in the dataset of the most frequent label. Accordingly, we conclude that to gain trust in persuasion, the quantity of truth revealed must always be greater than the quantity of lies told. We also conclude that the minimum number of signals used by a sender to obtain the Stackelberg game value can never exceed 2 symbols.

B. Related work

The interaction we study in this paper is a form of *Signalling game* [2] where the sender with access to private information sends messages to the receiver (uninformed player) about the private information to maximize its own utility. Bayesian persuasion introduced in [3] is a popular form of signalling games where the players move sequentially. Cheap talk introduced in [4] is also a form of signalling game where the players move simultaneously with the catch that the utility of the players is independent of the signals used by the sender. Nash equilibrium [5] is a popular approach used to capture the optimal strategies of the players if the players made their move simultaneously. We use the concept of Stackelberg equilibrium strategy [1] in this paper since the players move sequentially. *Screening games* [2] are another form of sequential games where a receiver with no access to information first commits to a strategy before the informed sender. Screening games with settings similar to our model were studied in [6]–[9].

Our sender-receiver model can also be viewed as a *task-oriented communication problem* or a *semantic communication problem* [10], where the goal of communication using semantics is to persuade the receiver to pick an action preferred by the sender. The signalling policy declared by the sender conveys the proportion of lies and truth hidden in the messages sent by the sender. This presentation of information via the signals or messages sent by the sender convinces the receiver to believe some facts and then accordingly pick an action preferred by the sender. Settings where the players have fully misaligned objectives like the setting in [11] are called *zero-sum games*. *Adversarial classification* [12] which is closely related to our work is also a major area of research in Machine learning.

In our previous work [13], [14], we studied a setting similar to this paper, where unlike the binary classification problem in this paper, every data point had a unique true label. The objective of the receiver there was to correctly recover or identify each data point from the messages sent by the sender. Additionally, unlike the pessimistic sender in our previous work, we deal with an optimistic sender in this paper. The players in [13] were restricted to play only deterministic strategies while players played behavioral strategies in [14]. With these changes in the setting of this paper, we seek to understand how a strategic sender behaves when it can lie only about two labels and also quantify the minimum amount of truth it must reveal to meet its objective. This is the gap we seek to fill in this area. To our surprise we could not only quantify informativeness, we also could quantify the maximum amount of lies the sender can tell to persuade the receiver to trust the sender for the two special classes of utility functions.

The paper is organized in the following order. We formulate our problem in Section II. Section III is dedicated to the construction and introduction of the Linear program and proving the equivalence between the Linear program and the Stackelberg game. We study two special classes of utility functions in Section IV and quantify the informativeness of the sender in Section V. Section VI concludes the paper.

II. Problem formulation

Let $\mathcal{X}$ denote a dataset of size q and $\mathcal{Z} = \{a, b\}$ denote the set of possible true labels for the data points in $\mathcal{X}$. We assume that the data points from $\mathcal{X}$ are generated uniformly at random and observed by the sender. Let $f : \mathcal{X} \longrightarrow \mathcal{Z}$

denote a true label function which is known to both the players: a sender and a receiver. Let $\mathscr{U} : \mathcal{Z} \times \mathcal{X} \to \mathbb{R}$ denote the utility function of the sender. If the sender persuades the receiver to classify $x \in \mathcal{X}$ as $z \in \mathcal{Z}$ then it obtains utility $\mathscr{U}(z, x)$. Let $\mathcal{Y}$ denote the set of signals used by the sender for signalling the sender, where $|\mathcal{Y}| = |\mathcal{X}|$. If $\mathcal{P}(\bullet)$ denotes the set of probability distributions on '$\bullet$', then let $\mathcal{A}_S := \mathcal{P}(\mathcal{Y}|\mathcal{X})$ and $\mathcal{A}_R := \mathcal{P}(\mathcal{Z}|\mathcal{Y})$ denote the strategy sets of the sender and the receiver respectively. Thus, $\mathcal{A}_S$ and $\mathcal{A}_R$ are collections of *behavioral strategies* of the sender and the receiver, respectively. Suppose that the sender plays some $\pi \in \mathcal{A}_S$ and the receiver plays $\sigma \in \mathcal{A}_R$. Then a data point $x \in \mathcal{X}$ is observed as $y \in \mathcal{Y}$ by the receiver with probability $\pi(y|x)$ and the receiver classifies y as $z \in \mathcal{Z}$ with probability $\sigma(z|y)$. If a sender plays $\pi \in \mathcal{A}_S$ and a receiver plays $\sigma \in \mathcal{A}_R$, then let $R_S(\pi, \sigma)$ and $R_R(\pi, \sigma)$ denote the respective expected utility of the sender and the receiver, where

$$\begin{aligned} R_S(\pi, \sigma) &:= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \sum_{z \in \mathcal{Z}} \pi(y|x)\sigma(z|y)\mathscr{U}(z, x) \\ R_R(\pi, \sigma) &:= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \sum_{z \in \mathcal{Z}} \pi(y|x)\sigma(z|y)\mathbb{I}_{\{f(x)=z\}}, \end{aligned} \tag{1}$$

where $\mathbb{I}$ represents the indicator function. In this setting our sender commits to a signalling policy (strategy) π first and the receiver after observing the choice of the sender accordingly picks a classification policy (strategy) σ which maximizes $R_R(\pi, \sigma)$. Let

$$\mathscr{B}_R(\pi) := \{\sigma \in \mathcal{A}_R | \sigma \in \arg\max_{\sigma' \in \mathcal{A}_R} R_R(\pi, \sigma')\}. \tag{2}$$

denotes the collection of all such strategies. We shall refer to $\mathscr{B}_R(\pi)$ as the *best response set* of the receiver for the strategy π of the sender. The sender is aware that for every choice of $\pi \in \mathcal{A}_S$, the receiver will always pick a strategy from $\mathscr{B}_R(\pi)$. Since the interaction is clearly a leader-follower game, we use the concept of *Stackelberg equilibrium strategy* of the sender to characterize the optimal strategies of the sender. The *Stackelberg game value* (SGV), where

$$SGV := \max_{\pi \in \mathcal{A}_S} \max_{\sigma \in \mathscr{B}_R(\pi)} R_S(\pi, \sigma) \tag{3}$$

is the maximum expected utility attainable by a sender in this game when a receiver only plays strategies from its best response set.

Definition 2.1: (Stackelberg equilibrium strategy) A strategy $\pi^* \in \mathcal{A}_S$ is called a Stackelberg equilibrium strategy of the sender if there exists a $\sigma^* \in \mathscr{B}_R(\pi^*)$ such that

$$R_S(\pi^*, \sigma^*) = SGV. \tag{4}$$

Clearly, $\max_{\sigma \in \mathscr{B}_R(\pi)} R_S(\pi, \sigma)$ is the *best-case expected utility* of the sender when a sender plays π. Therefore, every strategy of the sender whose best-case expected utility is exactly equal to the SGV must be a Stackelberg equilibrium strategy of the sender.

The main goal of our paper is to quantify the SGV for any given pair of $\mathscr{U}$ and f and identify the Stackelberg equilibrium strategies. We also seek to quantify the minimum number of data points that can be correctly classified at equilibrium. We achieve these goals by studying two linear programs which we cover in Section III and Section V.

III. Linear program for computing SGV

Note that for computing the SGV, one must compute the best-case expected utility for every strategy in the strategy space of the sender. However, the strategy space of the sender is infinite since it is a collection of probability distributions. To simplify the computation of the SGV, we introduce a linear program, which eliminates the dependency on the signals used. In the process of proving the equality of the optimal value of the linear program and the SGV, we also prove the equivalence of *optimal policies or solution* of the LP and the Stackelberg equilibrium strategies. We first present a few additional results which we use to prove our main theorem.

A. Best response of the receiver

A strategy of the receiver $\sigma \in \mathcal{A}_R$ is called a *deterministic strategy* of the receiver if $\sigma(z|y) = 1$ or $0, \forall z \in \mathcal{Z}, \forall y \in \mathcal{Y}$. We first prove the existence of a deterministic best response strategy of the receiver for every strategy of the sender. Suppose that the sender plays a strategy $\pi \in \mathcal{A}_S$. In the following lemma we show that for every observed y under π, it is optimal for the receiver to map y to some $z \in \mathcal{Z}$ which ensures that the sum of the probability of observation of those data points whose true label is z is the largest when the sender sends y while playing π.

Lemma 3.1: For any $\pi \in \mathcal{A}_S$, there exists a deterministic $\sigma \in \mathscr{B}_R(\pi)$ such that

$$\operatorname{supp}(\sigma(\bullet|y)) \equiv \arg\max_{z \in \mathcal{Z}} \sum_{x \in \mathcal{X}} \pi(y|x)\mathbb{I}_{\{f(x)=z\}}. \tag{5}$$

Proof: Fix a $\pi \in \mathcal{A}_S$. Recall that if $\sigma \in \mathscr{B}_R(\pi)$, then $\sigma \in \arg\max_{\sigma' \in \mathcal{A}_R} R_R(\pi, \sigma')$, where

$$\begin{aligned} &\max_{\sigma' \in \mathcal{A}_R} R_R(\pi, \sigma') \\ &= \sum_{y \in \mathcal{Y}} \max_{\sigma'(\bullet|y)} \sum_{x \in \mathcal{X}} \sum_{z \in \mathcal{Z}} \pi(y|x)\sigma'(z|y)\mathbb{I}_{\{f(x)=z\}} \\ &= \sum_{y \in \mathcal{Y}} \max_{\sigma'(\bullet|y)} \sum_{z \in \mathcal{Z}} \Big[\sum_{x \in \mathcal{X}} \pi(y|x)\mathbb{I}_{\{f(x)=z\}} \Big] \sigma'(z|y). \end{aligned} \tag{6}$$

Notice that the right-hand side of the above equation is linear in σ'. Therefore, there exists a deterministic $\sigma \in \mathscr{B}_R(\pi)$ which proves our lemma. ∎

B. Probabilistic classifier for the linear program

Given any pair of π and $\sigma \in \mathscr{B}_R(\pi)$, we construct a probabilistic classifier $\mu \in \mathcal{P}(\mathcal{Z}|\mathcal{X})$, where

$$\mu(z|x) := \sum_{y \in \mathcal{Y}} \pi(y|x)\sigma(z|y), \forall z \in \mathcal{Z}, \forall x \in \mathcal{X}. \tag{7}$$

Notice that for a pair of $x \in \mathcal{X}$ and $z \in \mathcal{Z}$ if $\mu(z|x) > 0$ under some probabilistic classifier μ, then x is classified

as z with probability $\mu(z|x)$. Next we show that every μ constructed from a pair of π and $\sigma \in \mathscr{B}_R(\pi)$ always satisfy the *trust constraints* defined below:

$$\sum_{x\in\mathcal{X}} \mu(z|x)\mathbb{I}_{\{f(x)=z\}} \geq \sum_{x'\in\mathcal{X}} \mu(z|x')\mathbb{I}_{\{f(x')\neq z\}}, \forall z \in \mathcal{Z}. \tag{8}$$

The trust constraints imply that under any μ constructed from a pair of π and σ, if you want some proportion of data points to be misclassified as z, then at least that proportion of data points whose true labels are z must be correctly classified. This is something we saw in our email filtering example introduced in Section I.

Lemma 3.2: Every μ constructed from a pair of π and $\sigma \in \mathscr{B}_R(\pi)$ using (7) must satisfy the trust constraints.

Proof: Suppose that μ is constructed from a pair of $\pi \in \mathcal{A}_S$ and $\sigma \in \mathscr{B}_R(\pi)$ using (7). For any $z \in \mathcal{Z}$, $\sum_{x\in\mathcal{X}} \mu(z|x)\mathbb{I}_{\{f(x)=z\}} = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} \pi(y|x)\sigma(z|y)\mathbb{I}_{\{f(x)=z\}}$. Since $\sigma \in \mathscr{B}_R(\pi)$, using Lemma 3.1, we get

$$\begin{aligned}
&\sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} \pi(y|x)\sigma(z|y)\mathbb{I}_{\{f(x)=z\}} \\
&= \sum_{y\in\mathcal{Y}} \sigma(z|y) \sum_{x\in\mathcal{X}} \pi(y|x)\mathbb{I}_{\{f(x)=z\}}, \\
&\geq \sum_{y\in\mathcal{Y}} \sigma(z|y) \sum_{x'\in\mathcal{X}} \pi(y|x')\mathbb{I}_{\{f(x')\neq z\}} \\
&= \sum_{x'\in\mathcal{X}} \mu(z|x')\mathbb{I}_{\{f(x')\neq z\}}.
\end{aligned} \tag{9}$$

This is true for any arbitrary μ constructed from a pair of π and $\sigma \in \mathscr{B}_R(\pi)$. ∎

The sender obtains expected utility $W(\mu)$ when a classifier μ is used. For any $\mu \in \mathcal{P}(\mathcal{Z}|\mathcal{X})$, let

$$W(\mu) := \sum_{x}\sum_{z\in\mathcal{Z}} \mu(z|x)\mathscr{U}(z,x) \tag{10}$$

$$\mathcal{C} := \{\mu \in \mathcal{P}(\mathcal{Z}|\mathcal{X})| \quad \mu \quad \text{satisfies} \quad (8)\}. \tag{11}$$

Thus, given any pair of $\mathscr{U}$ and f, we introduce a LP of the following form:

$$\begin{aligned}
L_{\mathscr{U};f}: \quad &\max_{\mu} W(\mu) \\
&\text{s.t.} \quad \mu \in \mathcal{C}.
\end{aligned} \tag{12}$$

In the following section, we show that the optimal value of $L_{\mathscr{U};f}$, denoted by $\mathrm{OPT}(L_{\mathscr{U};f})$ is exactly equal to the SGV.

C. Equality of Stackelberg game value and the optimal value of the LP

This section is dedicated to proving our main result, where we show that the optimal value of the LP constructed for a pair of μ and f is exactly equal to the SGV. We prove this by showing that every optimal solution of the LP is equivalent to some Stackelberg equilibrium strategy of the sender which can be constructed using (15). Additionally, we also show how we can construct an optimal solution of the LP by using a pair of Stackelberg equilibrium strategy of the sender and the receiver. Let $\mathrm{OPT}(L_{\mathscr{U}})$ denote the optimal value of the LP $L_{\mathscr{U};f}$.

Theorem 3.3: Given any pair of $\mathscr{U}$ and f,

$$\mathrm{OPT}(L_{\mathscr{U};f}) = SGV. \tag{13}$$

Proof: From (7), we know that given any pair of $\pi \in \mathcal{A}_S$ and $\sigma \in \mathscr{B}_R(\pi)$, we can construct a $\mu \in \mathcal{P}(\mathcal{Z}|\mathcal{X})$ such that $\mu \in \mathcal{C}$. Therefore,

$$SGV \leq \mathrm{OPT}(L_{\mathscr{U};f}). \tag{14}$$

Next we show that given any optimal solution μ^*, we can construct a pair of $\pi^* \in \mathcal{A}_S$ and $\sigma^* \in \mathscr{B}_R(\pi^*)$ such that $R_S(\pi^*, \sigma^*) = W(\mu^*)$. This will ensure that $R_S(\pi^*, \sigma^*) = SGV$, which will prove our theorem.

Fix a μ^*, where $\mu^*$is an optimal solution of the $L_{\mathscr{U};f}$. For every $z \in \mathcal{Z}$ such that $\mu^*(z|x) > 0$ for some $x \in \mathcal{X}$, pick $y_z \in \mathcal{Y}$ to construct the following $\pi \in \mathcal{A}_S$:

$$\pi(y_z|x) = \mu^*(z|x), \forall x \in \mathcal{X}, \forall z \in \mathcal{Z}. \tag{15}$$

This definition ensures that $\mathcal{Y}(\pi) = \{y_z \in \mathcal{Y} | \exists x \in \mathcal{X}, \exists z \in \mathcal{Z} \quad \text{s.t.} \quad \mu^*(z|x) > 0\}$. Recall that every feasible solution of $L_{\mathscr{U};f}$ must satisfy the trust constraints. Accordingly, we have

$$\sum_{x\in\mathcal{X}} \mu^*(z|x)\mathbb{I}_{\{f(x)=z\}} \geq \sum_{x'\in\mathcal{X}} \mu^*(z|x')\mathbb{I}_{\{f(x')\neq z\}}, \forall z \in \mathcal{Z}$$

which gives us

$$\sum_{x\in\mathcal{X}} \pi(y_z|x)\mathbb{I}_{\{f(x)=z\}} \geq \sum_{x'\in\mathcal{X}} \pi(y_z|x')\mathbb{I}_{\{f(x')\neq z\}}, \forall z \in \mathcal{Z}.$$

It is easy to see that $\sigma(z|y_z) = 1, \forall y_z \in \mathcal{Y}(\pi)$ is a best response strategy of the sender for the strategy π. But

$$\begin{aligned}
R_S(\pi,\sigma) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\sum_{z\in\mathcal{Z}} \pi(y|x)\sigma(z|y)\mathscr{U}(z,x) \\
&= \sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} \pi(y_z|x)\mathscr{U}(z,x) \\
&= \sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} \mu^*(z|x)\mathscr{U}(z,x) = W(\mu^*).
\end{aligned} \tag{16}$$

Therefore, for every optimal solution μ^* there exists a $\pi \in \mathcal{A}_S$ and a $\sigma \in \mathscr{B}_R(\pi)$ such that $R_S(\pi,\sigma) = W(\mu^*) = OPT(L_{\mathscr{U};f})$. Consequently, using (14) we conclude that (13) holds. ∎

D. Minimum signals required for obtaining SGV

The problem formulation ensures that the sender is free to use as many signals as it wants from the signal space $\mathcal{Y}$. But the SGV decides if the sender should use more than one signal or not for obtaining the SGV as the expected utility. Let $\mathcal{S}(\mathscr{U};f)$ quantify the minimum number of signals that is required by the sender to obtain SGV. Thus,

$$\mathcal{S}(\mathscr{U};f) = \min_{\pi\in\mathcal{A}_S^*:} |\mathcal{Y}(\pi)|, \tag{17}$$

where $\mathcal{A}_S^*$ is the collection of all Stackelberg equilibrium strategies of the sender. Let $P_O(\mathscr{U};f)$ denote the collection of all optimal solutions of $L_{\mathscr{U};f}$, where

$$P_O(\mathscr{U};f)=\{\mu^*\in\mathcal{C}|\mu^* \quad \text{is an opt. solution of} \quad L_{\mathscr{U};f}\}. \tag{18}$$

We show that if $(\mathscr{U};f)\in C_1$ then $\mathcal{S}(\mathscr{U};f)=1$ and if $(\mathscr{U};f)\in C_2$ then $\mathcal{S}(\mathscr{U};f)=2$, where

$$\begin{aligned} C_1 :=&\{(\mathscr{U};f):\exists\ z\in\mathcal{Z} \quad \text{and} \quad \exists\ \mu^*\in P_O(\mathscr{U};f) \quad \text{s.t.} \\ &\mu^*(z|x)=1,\forall x\in\mathcal{X}\} \quad \text{and} \\ C_2 :=&\{(\mathscr{U};f):\forall\mu^*\in P_O(\mathscr{U};f), \quad \exists \quad x_1,x_2\in\mathcal{X} \quad \text{s.t.} \\ &\mu^*(a|x_1),\mu^*(b|x_2)>0\}. \end{aligned} \tag{19}$$

Corollary 3.4: For any $(\mathscr{U};f)$,

$$\mathcal{S}(\mathscr{U};f)=\begin{cases}1 & \text{if} \quad (\mathscr{U};f)\in C_1\\ 2 & \text{if} \quad (\mathscr{U};f)\in C_2.\end{cases} \tag{20}$$

Proof: Let μ^* be any optimal solution of $L_{\mathscr{U};f}$. If for a fixed $z\in\mathcal{Z}$, $\mu^*(z|x)=0,\forall x\in\mathcal{X}$, then the corresponding π constructed in (15) will have $\mathcal{Y}(\pi)=\{y\}$ where $\pi(y|x)=1,\forall x\in\mathcal{X}$. As demonstrated in our main theorem proof, there exists a $\sigma\in\mathscr{B}_R(\pi)$ such that $R_S(\pi,\sigma)=W(\mu^*)=SGV$. Therefore, if $(\mathscr{U};f)\in C_1$, then the sender obtains utility exactly equal to SGV by using only one signal to for all data-points.

In a similar manner, for C_2 it is easy to see that from the construction of π for any optimal solution μ^* in (15), $|\mathcal{Y}(\pi)|=2$. Recall the equivalence relationship between an optimal policy of $L_{\mathscr{U};f}$ and some Stackelberg equilibrium strategy of the sender. Thus, if $(\mathscr{U};f)\in C_2$, then there exists no strategy $\pi\in\mathcal{A}_S^*$ with $|\mathcal{Y}(\pi)|=1$. This proves that with a minimal use of two signals by the sender, the sender obtains SGV if $(\mathscr{U};f)\in C_2$. ∎

With this result we can conclude that the minimum number of signals used by a sender to obtain the SGV can never exceed 2 symbols.

E. The case of multi-class classification

If we increase the size of the label class $\mathcal{Z}$ from 2 to any other natural number, the problem then transforms into a trust and persuasion problem for multi-class classification. The SGV for this multi-class classification problem can also be computed by solving a similar LP with $|\mathcal{Z}|$ trust constraints. In a similar manner, we can also conclude that $\mathcal{S}(\mathscr{U};f)$ lies between 1 and $|\mathcal{Z}|$, where the exact value is determined by the minimum size of the support of the optimal policies of $L_{\mathscr{U};f}$.

IV. Special class of utility functions

In this section, we study two special classes of utility functions. In our analysis, we draw an important conclusion on the maximum number of data points that can be misclassified by the receiver under any feasible policy or solution of $L_{\mathscr{U};f}$. We introduce the concept of *obscureness* of a true label function below to capture this quantity.

Definition 4.1: (Obscureness) Obscureness of the sender under a true label function f denoted by $\mathcal{O}(f)$ is defined as

$$\mathcal{O}(f):=\max_{\pi\in\mathcal{A}_S,\sigma\in\mathscr{B}_R(\pi)}\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}:f(x)\neq z}\sum_{y\in\mathcal{Y}}\pi(y|x)\sigma(z|y). \tag{21}$$

Thus, obscureness quantifies the maximum number of data points that can be misclassified when the receiver plays only best response to every sender strategy. Consequently, this is the maximum number of symbols that can be misclassified under every feasible policy in $L_{\mathscr{U};f}$.

Let $Q_f(a):=\sum_{x\in\mathcal{X}}\mathbb{I}_{\{f(x)=a\}}$ and $Q_f(b):=\sum_{x\in\mathcal{X}}\mathbb{I}_{\{f(x)=b\}}$. In our analysis, we actually prove that the obscureness of a true label function f is actually equal to the least frequent label under f denoted by Q_f, where

$$Q_f:=\min\{Q_f(a),Q_f(b)\}. \tag{22}$$

A. Strict preference of one class over other

In this section we study a special class of utility functions $\mathscr{U}\in\mathfrak{U}_1$, where the sender wants every data point whose true label is a to be misclassified by the receiver as b. Let $\mathfrak{U}_1$ be defined in the following way:

$$\begin{aligned}\mathfrak{U}_1 :=&\{\mathscr{U}:\mathcal{Z}\longrightarrow\mathcal{X}|\mathscr{U}(b,x)=1,\mathscr{U}(a,x)=0\\ &\forall x\in\mathcal{X} \quad \text{s.t.} \quad f(x)=a;\mathscr{U}(z,x)=0,\forall z\in\mathcal{Z},\\ &\forall x\in\mathcal{X} \quad \text{s.t.} \quad f(x)=b\}\end{aligned} \tag{23}$$

Thus, the sender gets utility 0 for every data point if it gets classified as a. We show that for every $\mathscr{U}\in\mathfrak{U}_1$, the optimal value of $L_{\mathscr{U};f}$ is Q_f. In the process of proving this, we also prove that the *obscureness* value for a pair of $\mathscr{U}\in\mathfrak{U}_1$ and f is exactly equal to Q_f.

Theorem 4.1: Given any pair of $\mathscr{U}\in\mathfrak{U}_1$ and f,

$$\mathrm{OPT}(L_{\mathscr{U};f})=Q_f. \tag{24}$$

Proof: There are three possible cases which arise when we compare $Q_f(a)$ and $Q_f(b)$:

Case 1: $Q_f(a)<Q_f(b)$

Note that $Q_f(a)$ number of 1's are the only positive utilities that the sender can obtain when all 'a' labelled data points are classified as label b. Thus, $\mathrm{OPT}(L_{\mathscr{U};f})\leq Q_f(a)=Q_f$. Consequently, we want a $\mu\in\mathcal{C}$ such that

$$\mathrm{OPT}(L_{\mathscr{U};f})=\sum_{x\in\mathcal{X}:f(x)=a}\mathscr{U}(b,x)=Q_f(a)=Q_f. \tag{25}$$

But we can construct a μ where $\mu(b|x)=1,\forall x$. This will ensure that μ satisfies (25) and

$$\begin{aligned}\sum_{x\in\mathcal{X}:f(x)=b}\mu(b|x)=Q_f(b)>\sum_{x\in\mathcal{X}:f(x)=a}\mu(b|x)=Q_f(a),\\ \text{and} \quad \sum_{x\in\mathcal{X}:f(x)=a}\mu(a|x)=0=\sum_{x\in\mathcal{X}:f(x)=b}\mu(a|x).\end{aligned} \tag{26}$$

Hence, μ is a feasible solution such that $W(\mu)=Q_f$. Therefore, $\mathrm{OPT}(L_{\mathscr{U};f})=Q_f$.

Case 2: $Q_f(a) = Q_f(b)$
The μ constructed in the previous case is a feasible and an optimal solution in this case as well since the trust constraints are satisfied by μ with equality for every $z \in \mathcal{Z}$.

Case 3: $Q_f(a) > Q_f(b)$
Notice that for a feasible μ, one must have $\sum_{x:f(x)=b} \mu(b|x) \geq \sum_{x:f(x)=a} \mu(b|x)$. But $Q_f(a) > Q_f(b)$ ensures that at max $\sum_{x:f(x)=b} \mu(b|x)$ can be equal to $Q_f(b)$. Thus, the trust constraints ensure that there does not exist any μ which gives expected utility value greater than $Q_f = Q_f(b)$. Therefore, proceeding in a similar way as demonstrated in the first case, we can show that a maximum of $Q_f = Q_f(b)$ number of data points can be misclassified. Thus, Q_f is the optimal value of $L_{\mathscr{U};f}$.

■

In all three cases of the proof, we saw that the maximum number of data points that can be misclassified under any feasible policy is given by Q_f. Therefore, irrespective of a balanced or imbalanced dataset, the obscureness of the sender is always equal to Q_f if $\mathscr{U} \in \mathfrak{U}_1$, resulting in the following corollary.

Corollary 4.2: Given any pair of $\mathscr{U} \in \mathfrak{U}_1$ and f, the obscureness value of the sender is exactly equal to Q_f.

Example 4.1. Consider a utility function $\mathscr{U} \in \mathfrak{U}_1$, where

$$\mathscr{U} = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{pmatrix} \end{array} \begin{matrix} a \\ b \end{matrix}.$$

Clearly $Q_f(a) = 3 > Q_f(b) = 2$. For this imbalanced dataset, more than two data points can never be misclassified as b since $Q_f(b) = 2$. Consequently, there exists no $\mu^* \in \mathcal{C}$ which can give $\sum_{x\in\mathcal{X}} \mu^*(b|x)\mathbb{I}_{\{f(x)=b\}} > 2$. Therefore, $\mathrm{OPT}(L_{\mathscr{U};f}) = \sum_{x\in\mathcal{X}} \mu^*(b|x)\mathbb{I}_{\{f(x)=b\}} = 2$, where μ^* is an optimal solution of $L_{\mathscr{U};f}$. □

B. Order based preference

Let $\mathfrak{U}_2$ denote the class of utility functions where $\mathcal{X} := \{x_1, x_2, \ldots, x_q\}$. Let $z \in \mathcal{Z}$ and $x \in \mathcal{X}$ be two arbitrary elements. If $\mathscr{U} \in \mathfrak{U}_2$ then

$$\mathscr{U}(z,x) := \begin{cases} 0 & f(x) = z \\ \mathscr{U}(z,x) > 0 & f(x) \neq z \end{cases}$$

and

$$\mathscr{U}(z,x_i) > \mathscr{U}(z,x_j), \forall j < i, \forall x_i, x_j \in \mathcal{X}.$$

This structure ensures that the sender would always prefer incorrect recovery of data points with indices close to q instead of 1. Let $A := \{\mathscr{U}(z,x) | z \in \mathcal{Z}, x \in \mathcal{X}\}$ and $B := \{A' \subseteq A | \ |A'| = Q_f\}$. We show that Q_f is obscureness value of the sender and the optimal value of $L_{\mathscr{U};f}$ for every $\mathscr{U} \in \mathfrak{U}_2$ is exactly equal to the sum of the largest Q_f utility values of the sender.

Theorem 4.3: Given any pair of $\mathscr{U} \in \mathfrak{U}_2$ and f,

$$\mathrm{OPT}(L_{\mathscr{U};f}) = \max_{A'\in B} \sum_{u\in A'} u \quad \text{and} \quad \mathcal{O}(f) = Q_f. \tag{27}$$

Proof: The structure of $\mathfrak{U}_2$ ensures that the sender gets positive utility for a data point if it gets misclassified by the receiver. Therefore, it needs to maximize the total number of data points that can be misclassified. But the total proportions of data points that can be misclassified as some $z \in \mathcal{Z}$ must be upper bounded by the total number of data points correctly classified as z. This follows from the trust constraints. Proceeding similarly, as demonstrated in the previous theorem for $\mathfrak{U}_1$, it is easy to see the obscureness of the sender is Q_f. Since Q_f is the maximum number of misclassified points under any feasible solution, the $\mathrm{OPT}(L_{\mathscr{U};f})$ must be given by the largest possible sum of Q_f utility values. This proves (27). ■

Example 4.2. Consider a utility function $\mathscr{U} \in \mathfrak{U}_2$, where

$$\mathscr{U} = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{matrix} \\ \begin{pmatrix} 1 & 2 & 0 & 0 & 3 & 4 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 & 3 & 4 & 7 \end{pmatrix} \end{array} \begin{matrix} a \\ b \end{matrix}.$$

Notice that $Q_f(b) = 4 < Q_f(a) = 5$. Using Theorem 4.3 we can conclude that $Q_f = 4$ data points can get misclassified which will get the sender the expected utility value exactly equal to $\mathrm{OPT}(L_{\mathscr{U};f})$. Notice that $\widehat{A} = \{\mathscr{U}(a,6) = 4, \mathscr{U}(b,7) = 3, \mathscr{U}(b,8) = 4, \mathscr{U}(b,9) = 7\} \in \arg\max_{A'\in B} \sum_{u\in A'} u$, where $\sum_{u\in\widehat{A}} u = 18$. There exists a feasible optimal policy μ^* for $L_{\mathscr{U};f}$, where $\mu^*(b|x) = 1, \forall x \in \{1,2,5,7,8,9\}$ and $\mu^*(a|x) = 1, \forall x \in \{3,4,6\}$, which satisfies the trust constraints since

$$\begin{aligned} \sum_{x\in 1,2,5} \mu^*(b|x) = 3 = \sum_{x\in\{7,8,9\}} \mu^*(b|x); \\ \sum_{x\in\{3,4\}} \mu^*(a|x) = 2 > \sum_{x\in 6} \mu^*(a|x) = 1. \end{aligned} \tag{28}$$

Therefore, $W(\mu^*) = 18 = \mathrm{OPT}(L_{\mathscr{U};f})$. □

This example clearly illustrates that if Q_f is the value of obscureness then the sender can never get the receiver to misclassify more than Q_f number of data points.

V. Informativeness of the Sender

A. The notion of informativeness

Our previous work involved studying the interaction between a sender and a receiver where the sender's utility function is given by $\mathscr{U} : \mathcal{X} \times \mathcal{X}$ and the true label function $f : \mathcal{X} \longrightarrow \mathcal{X}$ is the identity mapping. The sender wanted to maximize its utility while the receiver aimed to maximize the total number of data points correctly recovered. We showed that the Stackelberg equilibrium strategies in the

setting in [13] can be characterized by the *sender graph* we introduced in the paper. We also introduced the notion of *informativeness* in [13] to capture the minimum number of data points correctly recovered by the receiver at equilibrium. We showed that the informativeness of a utility function of the sender is exactly equal to the *vertex clique cover number* [16] of the sender graph. Since we quantified informativeness in our previous work, we also quantify the informativeness of a pair of $\mathscr{U}$ and f for this setting where the players play behavioral strategies. Earlier informativeness was dependent on the utility function of the sender, but in this particular setting, informativeness is dependent not only on the utility function but also on the true label function. This is because the optimal policies are dependent on the utility function of the sender and the true label function.

Definition 5.1 (Informativeness of the sender): If $\mathcal{A}_S^*$ denotes the collection of all Stackelebrg equilibrium strategies of the sender, then given any pair of $\mathscr{U}$ and f, the informativeness of the sender denoted by $\mathcal{I}(\mathscr{U};f)$ is

$$\mathcal{I}(\mathscr{U};f) := \min_{\pi^* \in \mathcal{A}_S^*} R_R(\pi^*, \sigma^*), \quad \text{where} \quad \sigma^* \in \mathscr{B}_R(\pi^*). \tag{29}$$

Lemma 5.1: Given any pair of $\mathscr{U}$ and f,

$$\mathcal{I}(\mathscr{U};f) := \min_{\mu^* \in P_O} \sum_{x \in \mathcal{X}} \mu^*(f(x)|x). \tag{30}$$

Proof: In our main theorem, we have shown that $SGV = \mathrm{OPT}(L_{\mathscr{U};f})$. Using the construction formula in (7), it is easy to see that for every $\pi^* \in S$, there exists a $\mu^* \in P_O$ such that

$$\sum_{x \in \mathcal{X}} \mu^*(f(x)|x) = R_R(\pi^*, \sigma^*), \forall \sigma^* \in \mathscr{B}_R(\pi^*). \tag{31}$$

Additionally, given any $\mu^* \in P_O$, we have also shown that we can construct a Stackelberg equilibrium strategy $\pi^* \in S$ such that (31) holds. Consequently, (30) follows immediately. ∎

Next, we show that the informativeness of the sender can also be computed by solving another Linear program, which we introduce next. Let $D(L_{\mathscr{U};f})$ denote the dual of $L_{\mathscr{U};f}$:

$$\begin{aligned} D(L_{\mathscr{U};f}): \quad & \min_{\gamma} \widehat{W}(\gamma) \\ & \text{s.t.} \quad \gamma \in F. \end{aligned} \tag{32}$$

Using this we can define a new LP whose optimal value is the informativeness of the sender. Let $I_{\mathscr{U};f}$ denote this new LP, where

$$\begin{aligned} I_{\mathscr{U};f}: \quad & \min_{\gamma,\mu} \sum_{x \in \mathcal{X}} \mu(f(x)|x) \\ & \text{s.t.} \quad \gamma \in F, \mu \in \mathcal{C} && (33) \\ & \quad W(\mu) = \widehat{W}(\gamma). && (34) \end{aligned}$$

The constraint (34) ensures that (μ, γ) is a feasible solution of $I_{\mathscr{U};f}$ if and only if μ and γ are optimal solutions of $L_{\mathscr{U};f}$ and $D(L_{\mathscr{U};f})$, respectively. This proves the following theorem.

Theorem 5.2: (Informativeness as the optimal value of an LP) Given any pair of $\mathscr{U}$ and f,

$$\mathcal{I}(\mathscr{U};f) = \mathrm{OPT}(I_{\mathscr{U};f}). \tag{35}$$

B. Informativeness in $\mathfrak{U}_1$ and $\mathfrak{U}_2$

In this section, we give the closed form value of informativeness for any $\mathscr{U} \in \mathfrak{U}_1$ and $\mathscr{U} \in \mathfrak{U}_2$. Since different values of $Q_f(a)$ and $Q_f(b)$ lead to different optimal policies, the informativeness is also dependent on the relationship between these two values. For any given pair of $\mathscr{U}$ and f let

$$\widehat{Q_f} := \max\{Q_f(a), Q_f(b)\}. \tag{36}$$

We show that the informativeness of the sender with utility function in class $\mathfrak{U}_1$ or $\mathfrak{U}_2$ is always $\widehat{Q_f}$.

Proposition 5.3: Given any pair of $\mathscr{U} \in \mathfrak{U}_1$ and f

$$\mathcal{I}(\mathscr{U};f) = \widehat{Q_f}. \tag{37}$$

Proof:

1) $\widehat{Q_f} = Q_f(a) = Q_f(b)$: Recall Theorem 4.1, where we showed that given any utility function, Q_f data points are correctly classified and Q_f data points are misclassified under the unique optimal solution. Therefore, $\mathcal{I}(\mathscr{U};f) = Q_f = \widehat{Q_f}$.
2) $\widehat{Q_f} = Q_f(a) > Q_f(b)$: For the case where $Q_f(a) > Q_f(b)$ in Theorem 4.1, we showed that only $Q_f(b)$ data points are correctly misclassified under every optimal policy. That leaves us with $Q_f(a) - Q_f(b)$ number of data points whose true label is a. But there are no more data points left to map whose true label is b since the optimal policy must meet the trust constraint. Therefore, the remaining $Q_f(a) - Q_f(b)$ number of data points must also be correctly classified as a under every optimal policy. Consequently, we get $\mathcal{I}(\mathscr{U};f) = Q_f(b) + (Q_f(a) - Q_f(b)) = Q_f(a) = \widehat{Q_f}$.
3) $\widehat{Q_f} = Q_f(b) > Q_f(a)$: From Theorem 4.1, it is clear that for $Q_f(a) < Q_f(b)$, $Q_f = Q_f(a)$ data points are correctly classified as a and $Q_f = Q_f(a)$ data points are misclassified as a. As for the remaining $Q_f(b) - Q_f(a)$ number of data points, these data points will never be misclassified under an optimal policy since there are no more remaining data points whose true label is a to constraint the number of misclassified data points. Therefore, we can conclude that $\mathcal{I}(\mathscr{U};f) = Q_f(a) + Q_f(b) - Q_f(a) = Q_f(b) = \widehat{Q_f}$ number of data points are correctly classified under every optimal policy of $L_{\mathscr{U};f}$.

∎

Recall Theorem 4.3, where we showed that for any $\mathscr{U} \in \mathfrak{U}_2$, the maximum number of data points that can be correctly misclassified is Q_f. This suggests that under every optimal policy only Q_f data points can be misclassified which results in $q - Q_f$ data points getting correctly classified, resulting in the following proposition:

Proposition 5.4: If $\mathscr{U} \in \mathfrak{U}_2$, then

$$\mathcal{I}(\mathscr{U}; f) = \widehat{Q_f}. \tag{38}$$

Proof: The proof follows using similar arguments used in the previous proposition. ■

Example 5.1. One can also verify these two propositions by computing the informativeness for the two examples illustrated in Section 4. Recall the $\mathscr{U}$ in example 4.1, where $Q_f(a) = 3$ and $Q_f(b) = 2$. We showed that under every optimal policy, only two data points are misclassified for this $\mathscr{U}$. Therefore, under every optimal policy $3 = Q_f(b)$ data points are getting classified correctly, which verifies Proposition 5.3.

Next recall $\mathscr{U}$ in example 4.2, where $Q_f(b) = 4 < Q_f(b) = 5$. We showed that only 4 data points can be misclassified under every optimal policy which resulted in $5 = Q_f(b)$ data points getting correctly classified under every optimal policy. This verifies Proposition 5.4. □

VI. Conclusion

We have a sender with access to a large dataset and the information on the true labels of every data point in the dataset. The sender gets incentives to persuade the receiver to misclassify some data points in the dataset while the receiver's goal is to maximize the probability of correctly classifying every data point. We characterized the optimal strategies of the sender and also quantified the minimum information that needs to be revealed by the sender to maximize its expected utilities. With our study, we concluded that a strategic sender's quantity of lies is always constrained. A sender can strategically persuade the receiver to trust its signals to misclassify some proportion of data points only by revealing more truth about data points and their true labels.

Acknowledgements

This research was supported by the Trust Lab Grant 2023 of IIT Bombay Trust Lab. The authors also acknowledge the support of TIH Foundation for IoT & IoE.

References

[1] T Basar and GJ Olsder. Dynamic noncooperative game theory. (classics in applied mathematics 23). 1999.

[2] Eric Rasmusen. Games and information: an introduction to game theory. 2007.

[3] Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615, 2011.

[4] Vincent P Crawford and Joel Sobel. Strategic information transmission. *Econometrica: Journal of the Econometric Society*, pages 1431–1451, 1982.

[5] Samson Lasaulce and Hamidou Tembine. *Game theory and learning for wireless networks: fundamentals and applications*. Academic Press, 2011.

[6] Anuj S. Vora and Ankur A. Kulkarni. Shannon meets Myerson: Information extraction from a strategic sender. https://arxiv.org/abs/2006.10641, 2022.

[7] Anuj S Vora and Ankur A Kulkarni. Zero error strategic communication. In *2020 International Conference on Signal Processing and Communications (SPCOM)*, pages 1–5. IEEE, 2020.

[8] Anuj S Vora and Ankur A Kulkarni. Information extraction from a strategic sender over a noisy channel. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 354–359. IEEE, 2020.

[9] Manish K. Singh and Ankur A. Kulkarni. Strategic multi-class classification for non-uniform preferences and its relation to incentive compatibility. In *2022 Eighth Indian Control Conference (ICC)*, pages 182–187, 2022.

[10] Arsham Mostaani, Thang X Vu, Shree Krishna Sharma, Van-Dinh Nguyen, Qi Liao, and Symeon Chatzinotas. Task-oriented communication design in cyber-physical systems: A survey on theory and applications. *IEEE Access*, 10:133842–133868, 2022.

[11] Ke Sun, Samir M Perlaza, and Alain Jean-Marie. $2\times$ 2 zero-sum games with commitments and noisy observations. In *2023 IEEE International Symposium on Information Theory (ISIT)*, pages 2254–2259. IEEE, 2023.

[12] Chidubem Arachie and Bert Huang. A general framework for adversarial label learning. *Journal of Machine Learning Research*, 22(118):1–33, 2021.

[13] Reema Deori and Ankur A Kulkarni. Information revelation through signalling. *Systems & Control Letters*, 169:105378, 2022.

[14] Reema Deori and Ankur A. Kulkarni. Informativeness and trust in bayesian persuasion, 2024.

[15] Harold W Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.

[16] Eran Halperin. Improved approximation algorithms for the vertex cover problem in graphs and hypergraphs. *SIAM Journal on Computing*, 31(5):1608–1623, 2002.