2019

Fairness And Feedback In Learning And Games

Shahin Jabbari
University of Pennsylvania, shahin.jabbari@gmail.com

Fairness And Feedback In Learning And Games

Abstract

In this thesis, we study fairness and feedback effects in game theory and machine learning. In game theory and economics, financial or technological networks are analyzed for feedback effects. These studies analyze how the connectivity benefits or risk of contagious shocks affect the individual agents or the structure of the network formed by these rational agents. Towards this direction, in the first part of this thesis, we study a series of novel network formation games and analyze the structural properties of the equilibrium networks.

Feedback effects can also occur in machine learning problems such as reinforcement learning or sequential allocation problems where the decisions of an algorithm over time can change the resources or actions available to the algorithm in the future as well as the environment in which the algorithm is operating. In the second part of this thesis, we study the effect of these feedback loops and ways to prevent them while also ensuring that the algorithm's actions and allocations satisfy natural notions of fairness. In particular we are interested in quantifying the cost of imposing fairness on learning algorithms.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Computer and Information Science

First Advisor

Michael Kearns

Keywords

Algorithmic Fairness, Feedback Loops, Game Theory, Machine Learning, Network Formation Games

Subject Categories

Computer Sciences

FAIRNESS AND FEEDBACK IN LEARNING AND GAMES

Shahin Jabbari

A DISSERTATION

in

Computer and Information Science

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2019

Supervisor of Dissertation

Michael Kearns
Professor and National Center Chair
Computer and Information Science

Graduate Group Chairperson

Rajeev Alur, Professor
Computer and Information Science

Dissertation Committee
Aaron Roth, Class of 1940 Bicentennial Term Associate Professor, Computer and Information Science
Sanjeev Khanna, Henry Salvatori Professor, Computer and Information Science
Sampath Kannan, Henry Salvatori Professor, Computer and Information Science
Jamie Morgenstern, Assistant Professor, School of Computer Science at Georgia Tech

Acknowledgement

I would like to thank the members of my committee: Sampath Kannan, Sanjeev Khanna, Jamie Morgenstern and Aaron Roth. Also I would like to warmly thank my advisor, Michael Kearns, for his patience and guidance during the past few years.

ABSTRACT

FAIRNESS AND FEEDBACK IN LEARNING AND GAMES

Shahin Jabbari

Michael Kearns

In this thesis, we study fairness and feedback effects in game theory and machine learning. In game theory and economics, financial or technological networks are analyzed for feedback effects. These studies analyze how the connectivity benefits or risk of contagious shocks affect the individual agents or the structure of the network formed by these rational agents. Towards this direction, in the first part of this thesis, we study a series of novel network formation games and analyze the structural properties of the equilibrium networks.

Feedback effects can also occur in machine learning problems such as reinforcement learning or sequential allocation problems where the decisions of an algorithm over time can change the resources or actions available to the algorithm in the future as well as the environment in which the algorithm is operating. In the second part of this thesis, we study the effect of these feedback loops and ways to prevent them while also ensuring that the algorithm's actions and allocations satisfy natural notions of fairness. In particular we are interested in quantifying the cost of imposing fairness on learning algorithms.

Table of Contents

List of Tables

List of Illustrations

Chapter 1

Introduction

This thesis is devoted to the study of feedback effects both in game theory and machine learning. In the game theory chapters, we revisit classic network formation games modified to incorporate feedback effects by introducing an adversarial attack that can spread through the collectively formed network. In the machine learning chapters, we focus on classic learning settings where the learning is sequential and feedback loops might occur as a result of the interactive nature of the learning process. These feedback effects can in turn impose fairness concerns and we study these effects both in theory and practice.

In the first part of this thesis (Chapters 2 and 3) we study network formation games where strategic agents (or vertices) benefit from connectivity to other agents but also incur a cost for forming these connections. Most of the previous work on network formation games only focus on the case where the cost to an agent is direct (e.g. via purchasing a link). Motivated by scenarios as diverse as technological vulnerability or biological contagion we consider a game which incorporates indirect costs i.e. in our model an agent's connection might also expose her to the negative contagious shocks the network might endure after the formation.

In our first model of network formation games, in Chapter 2, we begin with the well-studied reachability network formation game [9] and modify this game by introducing an adversary that is allowed to examine the network and choose a single vertex to attack. The attack then spreads throughout the network destroying all the vertices in the connected component of the originally attacked vertex. The agents crucially have the option of purchasing immunization against the attack for a fixed cost. Thus the attack can only spread to the vertices that are reachable solely by unimmunized vertices from the originally attacked vertex.

We introduce several adversarial attacks including an adversary that seeks to maximize the destruction, an adversary whose goal is to minimize the social welfare and even an adversary

than randomly selects a vertex to attack. Our focus is to understand the properties of the formed networks in equilibria. In particular, we are interested in understanding the tension between the connectivity benefits and indirect risks of contagious shocks with questions such as "Would the introduction of attack and immunization result in over-building or erode in welfare in equilibrium networks compared to the attack-free game?"

Our first result shows that the edge density of the networks formed in the equilibria of the new game over n vertices is bounded by $2n-4$ and this upper bound is tight. This is less than twice as the number of edges in the equilibria of the original reachability network formation game. So the amount of overbuilding is sharply limited. In our second result, we show that in any non-trivial equilibrium network, the social welfare is at least $n^2 - o(n^{5/3})$, which is asymptotically the maximum welfare possible with a polynomial rate of convergence.

The model analyzed in Chapter 2 relies on the assumption that the attack spreads deterministically to adjacent vertices. However, in most real-world scenarios the spread of the attack is non-deterministic (e.g. the spread of a contagions disease over a biological network). This rather simple modification makes the game drastically complex as now even computing the utility of vertices becomes #P-complete. Regardless of this difficulty, in Chapter 3, we show that the results of Chapter 2 are robust to this modification. We focus on the regime where the cost of immunization is high and an adversary that selects a vertex uniformly at random to start the attack. We show that when the attack spreads with constant probability (independent of size of the network) and according to the independent cascade model [61], the edge density in the equilibrium networks over n vertices is bounded by $O(n \log(n))$, so again the amount of overbuilding compared to the original reachability game is limited. We also show that the in any non-trivial equilibrium network the social welfare is $\Omega(n^2)$ as long as the network is not *too dense*. Therefore, many of the properties of the game with deterministic spread are preserved in the new variant.

The second part of this thesis (Chapters 4-6) is regarding the recent and emerging area of machine learning referred to as *algorithmic fairness*. Algorithms have been used to make

decisions in seemingly *low impact* tasks like email spam filtering for a long time. With the advance of machine learning and the invention of sophisticated and accurate learning algorithms, machine learning has been recently applied to make *high impact* decisions like hiring [76], sentencing [10] and lending [18] which has been traditionally made by humans.

In addition to improving the accuracy of the decisions over a human decision maker it is tempting to think that the learning algorithms would not inherit the biases that are usually exhibited in human decision makers towards specific subgroups characterized e.g. by race or gender. However, recent empirical evidence suggests that this is not the case [85]. There are several ways to explain why the algorithms exhibit bias or behave *unfairly*. First, usually the learning algorithm is trained on historic data which itself can contain bias. It is then inevitable that the algorithm inherits these biases from the data. This feedback effect can be amplified with continuous utilization of the algorithm over time in a scenario where the data from the past interactions are repeatedly used to fine tune the algorithm's decisions for future. To make matters worse and more complicated, it has also been observed that the algorithm can exhibit unfairness even when the historic data is unbiased or when the process of learning starts with no initial data (e.g. in online learning [52]). In addition, the successes of machine learning algorithms are highly dependent on the amount of training data available to them. Lack of sufficient amount of training data for *minority* groups has also been credited as another reason behind the unfair behavior of the algorithms.

Before describing how to deal with unfairness, we note that there are several proposals and definitions for what it means for a learning algorithm to be fair (see e.g. [30, 45, 63, 78, 90]). Not only these definitions are drastically different and somewhat task-dependent, unless in trivial cases, satisfying some of these definitions at the same time is also impossible [37, 63]. This implies that different solutions should be designed to remedy different type of unfairness in different learning settings. Regardless of these differences, usually a fairness notion can be interpreted as a constraint imposed on the learning algorithm. Therefore, enforcing fairness would usually deter the prediction utility of the algorithm. Studying the trade-off

between fairness and utility is an important direction in the algorithmic fairness literature.

The approaches used in the algorithmic fairness literature can be roughly divided into three categories. In pre-processing [44, 53], before any learning is done, the training data itself would be modified so that the desired fairness notion is satisfied. After the pre-processing step, a black-box learning algorithm can be used to learn a predictor hoping that fairness would continue to hold when using this learned predictor on the modified data. In-processing [19, 55] can refer to two types of approaches. In the first type, initially a predictor is learned ignoring fairness altogether. Then the predictor is modified to account for fairness (e.g. by shifting the decision boundary in favor of the minority group). In the second type, fairness is usually written as a constraint in the optimization problem used to derive a predictor. Then the constrained optimization problem is solved either directly or using a heuristic. In post-processing [45, 88], first a prediction model is learned, again ignoring fairness altogether. Then the outcome of the predictor will be distorted only as a function of the group the data point belongs to and the prediction of the algorithm. Crucially, the modification in post-processing cannot depend on the group-agnostic attributes of the input and this is the essential difference between in-processing and post-processing.

Finally, most of the current body work on algorithmic fairness is on classification which is the most well-understood task in machine learning aiming to predict binary outcomes. We consider three different machine learning settings, introduce new notions of fairness for each setting, design algorithmic frameworks to satisfy these notions of fairness and finally quantify the trade-off between fairness and utility in each of these settings.

More specifically, in Chapter 4, we consider the reinforcement learning setting where unlike the stationary dynamics of the classification task has the property that the decisions made by the algorithm in the past can change the environment in which the algorithm is operating as well as the possibilities available to the algorithm in the future. Our notion of fairness in reinforcement learning is inspired by the weakly meritocratic notion of fairness introduced by Joseph et al. [52] for online learning. In reinforcement learning, our notion implies that

an algorithm cannot favor an action with a lower *true* long-term potential to an action with a higher *true* long-term potential. Since the true long-term potential of an action is unknown to the algorithm, the constraint of fairness mandates the learning algorithm to explore extensively. Given this observation, we show that imposing such a fairness constraint to an algorithm makes efficient learning impossible. We further design an efficient in-processing learning algorithm for a slightly relaxed version of this fairness notion.

In Chapter 5 we consider the regression setting where the goal is to predict a real valued number using a linear model. We introduce a rich family of fairness metrics for regression models that take the form of a fairness regularizer and apply them to the standard loss functions for linear and logistic regression. Our family of fairness metrics covers the spectrum from the type of group fairness that are common in the classification setting to much stronger notions of individual fairness. Our framework allows for computationally efficient in-processing algorithms and we empirically study the trade-off between fairness and the prediction utility of our learning algorithm across several real-world datasets where fairness is a concern. Our analysis reveals that the trade-off between fairness and utility can vary significantly when considering different notions of fairness or different learning domains.

Finally, in Chapter 6, we study allocation problems under a notion of fairness inspired by the general principle of equality of opportunity [45]. We assume a population is divided into groups and each group is consisted of candidates that the algorithm (or the allocator) would like to receive the resource and non-candidates which are the remaining members of the group. The goal is to maximize utility defined to be the number of candidates that receive the resource in the allocation of the algorithm. Our notion of fairness requires that conditioned on being a candidate for the resource, the probability of receiving the resource should be independent of the group. We further study the learning problem in our fairness constrained allocation problem where the algorithm needs to reason about the distributions of candidates in each group to ensure both fairness and optimal utility. We provide efficient in-processing algorithms for our setting and also study the trade-off between fairness and

utility of the allocation both in theory and empirically over a real-world dataset.

Bibliography Chapter 2 is based on Goyal et al. [42]. Chapter 3 is adapted from Chen et al. [22]. The results in Chapter 4 are from Jabbari et al. [48]. Chapter 5 is based on Berk et al. [12]. Finally, Chapter 6 is based on Elzayn et al. [31]. We refer the reader to these papers for omitted proofs and details. Additional publications while at University of Pennsylvania are as follows: Assadi et al. [8], Jabbari et al. [47], Berk et al. [13] and Dong et al. [28]. We thank K. Amin, S. Assadi, R. Berk, Y. Chen, J. Dong, H. Elzayn, S. Goyal, H. Heidari, J. Hsu, M. Joseph, C. Jung, M. Kearns, S. Khanna, J. Morgenstern, S. Neel, R. Rogers, A. Roth, Z. Schutzman and Z.S. Wu for their collaborations in these works.

Chapter 2

Network Formation Games with Attack and Deterministic Spread

2.1. Introduction

In network formation games, distributed and strategic agents receive some benefit from their connectedness to others, but also incur some cost for forming these links. Much research in this area [9, 16, 34] studies the structure of equilibrium networks formed as the result of various choices for the network benefit function, as well as the social welfare in equilibria. In many network formation games, the costs incurred from forming links are direct: each edge costs $C_{\mathrm{E}} > 0$ for an agent to purchase. Recently, motivated by scenarios as diverse as financial crises, terrorism and technological vulnerability, games with indirect connectivity costs have been considered: an agent's connections expose her to negative, contagious shocks the network might endure.

We begin with the simple and well-studied *reachability* network formation game [9], in which players purchase links to each other, and enjoy a network benefit equal to the size of their connected component in the collectively formed graph. We modify this model by introducing an adversary who is allowed to examine the network, and choose a single vertex or player to attack. This attack then spreads throughout the entire connected component of the originally attacked vertex, destroying all of these vertices. Crucially however, players also have the option of purchasing *immunization* against attack. Thus the attack spreads only to those non-immunized (or *vulnerable*) vertices reachable from the originally attacked vertex. We examine several natural adversarial attacks such as an adversary that seeks to maximize destruction, an adversary that randomly selects a vertex for the start of infection and an adversary that seeks to minimize the social welfare of the network post-attack to

name a few. A player's overall payoff is thus the expected size of her post-attack component, minus her edge and immunization expenditures.

Our game can be viewed as a stylized model for settings where reachability rather than centrality is the primary interest in joining a network vulnerable to adversarial attack. Examples include technological networks such as the Internet, where packet transmission times are sufficiently low that being "central" [34] or a "hub" [16] is less of a concern, but in the presence of attacks such as viruses or DDoS, mere reachability may be compromised. Parties may reduce risks via costly measures such as anti-virus. In a financial setting, vertices might represent banks and edges credit/debt agreements. The introduction of an attractive but extremely risky asset is a threat or attack on the network that naturally seeks its largest accessible market, but can be mitigated by individual institutions adopting balance sheet requirements or leverage restrictions. In a biological setting, vertices could represent humans, and edges physical proximity or contact. The attack could be an actual biological virus that randomly infects an individual and spreads by physical contact through the network; again, individuals may have the option of immunization. While our simplified model is obviously not directly applicable to any of these examples in detail, we do believe our results provide some high-level insights about the strategic tensions in such scenarios. See Section 2.8 for discussion of some variants of our model.

Immunization against attack has recently been studied in games played on a network where risk of contagious shocks are present [21] but only in the setting in which the network is first designed by a centralized party, after which agents make individual immunization decisions. We endogenize both these aspects, which leads to a model incomparable to this earlier work.

The original reachability game [9] permitted a sharp and simple characterization of all equilibrium networks: any tree as well as the empty graph. We demonstrate that once attack and immunization are introduced, the set of possible equilibria becomes considerably more complex, including networks that contain multiple cycles, as well as others which are disconnected but nonempty. This diversity of equilibrium topologies leads to our primary

questions of interest: How dense can equilibria become? In particular, does the presence of the attacker encourage the creation of massive redundancy of connectivity? Moreover, does the introduction of attack and immunization result in dramatically lower social welfare compared to the original game?

Our Results and Techniques The main theoretical contributions of this work are to show that our game still exhibits edge sparsity at equilibrium, and has high social welfare properties despite the presence of attacks. First we show that under a very mild assumption on the adversary's attack model, the equilibrium networks with $n \geq 4$ players have at most $2n - 4$ edges, fewer than twice as many edges as any nonempty equilibria of the original reachability game without attack. We prove this by introducing an abstract representation of the network and use tools from extremal graph theory to upper bound the resources globally invested by the players to mitigate connectivity disruptions due to any attack, obtaining our sparsity result.

We then show that with respect to several adversarial attack models, in any equilibrium with at least one edge and one immunized vertex, the resulting network is connected. These results imply that any *new* equilibrium network (i.e. one which was not an equilibrium of the original reachability game) is either a sparse but connected graph, or is a forest of unimmunized vertices. The latter occurs only in the rather unnatural case where the cost of immunization or edges grows with the population size, and in the former case we further show the social welfare is at least $n^2 - O(n^{5/3})$, which is asymptotically the maximum possible with a polynomial rate of convergence. These results provide us with a complete picture of social welfare in our model. We show the welfare lower bound by first proving any equilibrium network with both immunization and an edge is connected, then showing that there cannot be many targeted vertices who are *critical* for global connectivity, where critical is defined formally in terms of both the vertex's probability of attack and the size of the components remaining after the attack. Thus players myopically optimizing their own utility create highly resilient networks in presence of attack.

We complement our theory with simulations demonstrating fast and general convergence of *swapstable* best response, a type of limited best response which generalizes linkstable best response but is much more powerful in our game. The simulations provide a dynamic counterpart to our static equilibrium characterizations and illustrate a number of interesting further features of equilibria, such as heavy-tailed degree distributions.

Organization We formally present our model and review some related work in Section 2.2. In Section 2.3 we describe some interesting topologies that arise as equilibria in our model illustrating the richness of the solution space. We present our sparsity result and lower bound on welfare in Sections 2.4 and 2.5, respectively. Sections 2.6 and 2.7 describe our simulations and behavioral experiment, respectively. We conclude with some directions for future work in Section 2.8.

2.2. Model

We assume the n vertices of a graph (network) correspond to individual players. Each player has the choice to purchase edges to other players at a cost of $C_{\mathrm{E}} > 0$ per edge. Each player additionally decides whether to immunize herself at a cost of $C_{\mathrm{I}} > 0$ or remain *vulnerable*.

A (pure) *strategy* for player i (denoted by s_i) is a pair consisting of the subset of players i purchased an edge to and her immunization choice. Formally, we denote the subset of edges which i buys an edge to as $x_i \subseteq \{1, \ldots, n\}$, and the binary variable $y_i \in \{0, 1\}$ as her immunization choice ($y_i = 1$ when i immunizes). Then $s_i = (x_i, y_i)$. We assume that edge purchases are unilateral i.e. players do not need approval or reciprocation in order to purchase an edge to another but that the connectivity benefits and risks are bilateral. Furthermore, we restrict our attention to pure strategy equilibria and our results show they exist and are structurally diverse.

Let $\mathbf{s} = (s_1, \ldots, s_n)$ denote the strategy profile for all the players. Fixing $\mathbf{s}$, the set of edges purchased by all the players induces an undirected graph and the set of immunization

decisions forms a bipartition of the vertices. We denote a game *state* as a pair $(G, \mathcal{I})$, where $G = (V, E)$ is the graph induced by the edges purchased by the players and $\mathcal{I} \subseteq V$ is the set of players who decide to immunize. We use $\mathcal{U} = V \setminus \mathcal{I}$ to denote the vulnerable vertices i.e. the players who decide not to immunize. We refer to a subset of vertices of $\mathcal{U}$ as a *vulnerable region* if they form a maximally connected component. We denote the set of vulnerable regions by $\mathcal{V} = \{\mathcal{V}_1, \ldots, V_k\}$ where each $\mathcal{V}_i$ is a vulnerable region.

Fixing a game state $(G, \mathcal{I})$, the adversary inspects the formed network and the immunization pattern and chooses to attack some vertex. If the adversary attacks a vulnerable vertex $v \in \mathcal{U}$, then the attack starts at v and spreads, killing v and any other vulnerable vertices reachable from v. Immunized vertices act as "firewalls" through which the attack cannot spread. *In this work we restrict the adversary to only pick one seed to start the attack.*

More precisely, the adversary is specified by a function that defines a probability distribution over vulnerable regions. We refer to a vulnerable region with non-zero probability of attack as a *targeted region* and the vulnerable vertices inside of a targeted region as *targeted vertices*. We denote the targeted regions by $\mathcal{T} = \{\mathcal{T}_1, \ldots, \mathcal{T}_{k'}\}$ where each $\mathcal{T}' \in \mathcal{T}$ denotes a targeted region and $k' \leq k$.

$\mathcal{T} = \emptyset$ corresponds to the adversary making no attack, so player i's *utility* (or *payoff*) is equal to the size of her connected component minus her expenses (edge purchases and immunization). When $|\mathcal{T}| > 0$, then player's i expected utility (fixing a game state) is equal to the expected size of her connected component less her expenditures, where the expectation is taken over the adversary's choice of attack (a distribution on $\mathcal{T}$) and the size of the connected component of a vertex is defined to be zero in the event she is killed. Formally, let $\Pr[\mathcal{T}']$ denote the probability of attack to targeted region $\mathcal{T}'$ and $CC_i(\mathcal{T}')$ the size of the connected component of player i post-attack to $\mathcal{T}'$. Then the expected utility of

i in strategy profile s denoted by $u_i(\mathbf{s})$ is precisely

$$u_i(\mathbf{s}) = \sum_{\mathcal{T}' \in \mathcal{T}} \left(\Pr\left[\mathcal{T}'\right] CC_i\left(\mathcal{T}'\right) \right) - |x_i| C_{\mathrm{E}} - y_i C_{\mathrm{I}}.$$

We refer to the sum of expected utilities of all players playing $\mathbf{s}$ as the *(social) welfare* of $\mathbf{s}$.

Examples of Adversaries We highlight several natural adversaries that fit into our framework. We begin with an adversary whose goal is to maximize the number of agents killed.

Definition 1. *The* maximum carnage *adversary attacks the vulnerable region of maximum size. If there are multiple such regions, the adversary picks one of them uniformly at random. Once a targeted region is selected for the attack, the adversary selects a vertex inside of that region uniformly at random to start the attack.*

Then a targeted region with respect to a maximum carnage adversary is a vulnerable region of maximum size and the adversary defines a uniform distribution over such regions (see Figure 1). We now introduce another natural but less sophisticated adversary which starts an attack by picking a vulnerable vertex at random.

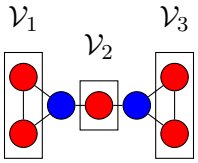


Figure 1: Blue and red vertices denote $\mathcal{I}$ and $\mathcal{U}$, respectively. The probability of attack to the vulnerable regions $\mathcal{V}_1, \mathcal{V}_2$ and $\mathcal{V}_3$ (in that order) for each adversary are as follows. maximum carnage: 0.5, 0, 0.5; random attack: 0.4, 0.2, 0.4; maximum disruption: 0, 1, 0.

Definition 2. *The* random attack *adversary attacks a vulnerable vertex uniformly at random.*

So every vulnerable vertex is targeted with respect to the random attack adversary and the adversary induces a distribution over targeted regions such that the probability of attack to a targeted region is proportional to its size (see Figure 1). Lastly, we define another natural adversary whose goal is to minimize the post-attack social welfare.

Definition 3. *The* maximum disruption *adversary attacks the vulnerable region which minimizes the post-attack social welfare. If there are multiple such regions, the adversary picks one of them uniformly at random. Once a targeted region is selected for the attack, the adversary selects a vertex inside of that region uniformly at random to start the attack.*

This adversary only attacks those vulnerable regions which minimize the post-attack welfare and the adversary defines a uniform distribution over such regions (again see Figure 1).

Equilibrium Concepts We analyze the networks formed in our game under two types of equilibria. We model each of the n players as strategic agents who choose deterministically which edges to purchase and whether or not to immunize, knowing the exogenous behavior of the adversary defined as above. A strategy profile $\mathbf{s}$ is a *pure-strategy Nash equilibrium* (Nash equilibrium for short) if, for any player i, fixing the behavior of the other players to be $\mathbf{s}_{-i}$, the expected utility for i cannot strictly increase playing any action s_i' over s_i.

In addition to Nash, we study another equilibrium concept that is closely related to linkstable equilibrium (see e.g. [15]), a bounded-rationality generalization of Nash. We refer to this concept as *swapstable equilibrium*. This equilibrium concept was first introduced by Lenzner [68] under the name *greedy equilibrium*. A strategy profile is a swapstable equilibrium if no individual agent's expected utility (fixing other agent's strategies) can strictly improve under any of the following *swap deviations:* (1) Dropping any single purchased edge, (2) Purchasing any single unpurchased edge, (3) Dropping any single purchased edge and purchasing any single unpurchased edge, (4) Making any one of the deviations above, and also changing the immunization status.

The first two deviations correspond to the standard linkstability. The third permits the more powerful *swapping* of one purchased edge for another. The last additionally allows reversing immunization status. Our interest in swapstable networks derives from the fact that while they only consider "simple" or "local" deviation rules, they share several properties with Nash networks that linkstable networks do not. In that sense, swapstability is a bounded

rationality concept that moves us closer to full Nash. Intuitively, in our game (and in many of our proofs), we exploit the fact that if a player is connected to some other set of vertices via an edge to a targeted vertex, and that set also contains an immune vertex, the player would prefer to connect to the immune vertex instead. This deviation involves a swap not just a single addition or deletion. It is worth mentioning explicitly that by definition every Nash equilibrium is a swapstable equilibrium and every swapstable equilibrium is a linkstable equilibrium. The reverse of none of these statements are true in our game. We also point out that the set of equilibrium networks with respect to adversaries defined in Definitions 1, 2 and 3 are disjoint.

2.2.1. Related Work

The problem of strategic network design and defense has been extensively studied in economics, electrical engineering, and computer science (see e.g. [4, 5, 41, 81]). Most of the existing work takes the network as given and examines optimal security choices (see e.g. [7, 26, 43, 58, 67]). To the best of our knowledge, our we offer the first model in which both links and defense (immunization) are chosen by the players.

Combining linking and immunization within a common framework yields new insights. We start with a discussion of the network formation literature. In a setting with no attack, our model reduces to the original model of one-sided reachability network formation of Bala and Goyal [9]. They showed that a Nash equilibrium network is either a tree or an empty network. By contrast, we show that in the presence of a security threat, Nash networks exhibit very different properties: both networks containing cycles and partially connected networks can emerge in equilibrium. Moreover, we show that while networks may contain cycles, they are sparse (we provide a tight upper bound on the number of links in any equilibrium network of our game).

Regarding security, Cerdeiro et al. [21] study optimal design of networks in a setting where players make immunization choices against a maximum carnage adversary but the network

design is given. A vertex v is defined to be k-critical in a connected network if the size of the largest connected component after removing v is k. Cerdeiro et al. [21] show that an optimal network is either a hub-spoke or a network containing k-critical vertices or a partially connected network (observe that a k-critical vertex can secure $n - k$ vertices by immunization). We extend this work by showing that there is a pressure toward the emergence of k-critical vertices even when linking is decentralized. We also contribute to the study of welfare costs of decentralization. Cerdeiro et al. [21] show that the Price of Anarchy (PoA) is bounded, when the network is centrally designed while immunization is decentralized (their welfare measure includes the edge expenditures of the planner). By contrast, we show that the PoA is unbounded when both decisions are decentralized. Although we also show that non-trivial equilibrium networks with respect to various adversaries have a PoA very near 1. This highlights the key role of linking and resonates with the original results on the PoA in for pure network formation games (see e.g. [40]).

Blume et al. [16] study network formation where new links generate direct (but not reachability) benefits, infection can flow through these links and immunization is not a choice. They demonstrate a fundamental tension between socially optimal and stable networks: the former lie just below a linking threshold that keeps contagion under check, while the latter admit linking just above this threshold, leading to extensive contagion and very low payoffs.

Furthermore, Kliemann [64] introduced a reachability network formation game with attacks but without defense. In their model, the attack also happens after the network is formed and the adversary destroys exactly one *link* in the network (with no spread) according to a probability distribution over links that can depend on the structure of the network. They show equilibrium networks in their model are chord-free and hence sparse. We also show an abstract representation of equilibrium networks in our model corresponds to chord-free graphs and then use this observation to prove sparsity. While both models lead to chord-free graphs in equilibria, the analysis of *why* these graphs are chord-free is quite different. In their model, the deletion of a single link destroys at most one path between any pair of

vertices. So if there were two edge-disjoint paths between any pairs of vertices, they will certainly remain connected after any attack. In our model the adversary attacks a vertex and the attack can spread and delete many links. This leads to a more delicate analysis. The welfare analysis is also quite different, since the deletion of an edge can cause a network to have at most two connected components, while the deletion of (one or more) vertices might lead to many connected components.

Finally, in a follow up work, Friedrich et al. [38] study the complexity of computing Nash best response for our game with respect to the maximum carnage and random adversaries.

2.3. Diversity of Equilibrium Networks

In contrast to the original reachability network formation game [9], our game exhibits equilibrium networks which contain cycles, as well as non-empty graphs which are not connected. Figure 2 gives several examples of specific Nash equilibrium networks with respect to the maximum carnage adversary for small populations, each of which is representative of a broad family of equilibria for large populations and a range of values for C_{E} and C_{I}. In this chapter, we represent immunized and vulnerable vertices as blue and red, respectively. Although we treat the networks as undirected graphs (since the connectivity benefits and risks are bilateral), we use directed edges in some of the figures to denote which player purchased the edge e.g. $i \to j$ means that i has purchased an edge to j. Finally, we use the maximum carnage adversary in many of our illustrations throughout this chapter because both the adversary's choice of attack and verifying certain properties are the easiest in this model compared to other natural models of Section 2.2. The examples in Figure 2 show that the tight characterization of the reachability game, where equilibrium networks are either empty graph or trees, fails to hold for our more general game (though such graphs can also form at equilibrium in our game). However, in the following sections, we show that an approximate version of this characterization continues to hold for several adversaries.

On the one hand, examples in Figure 2 show that equilibrium networks can be denser in our

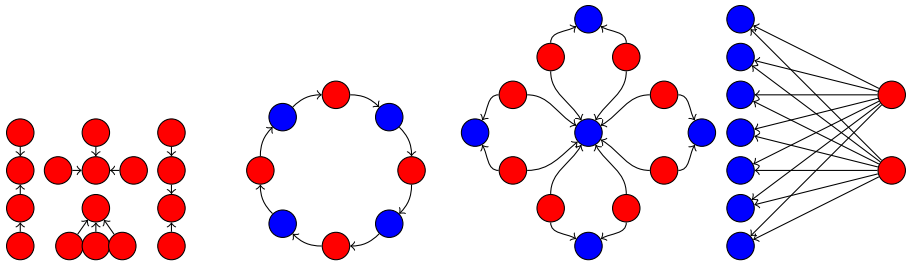


Figure 2: Examples of equilibria with respect to the maximum carnage adversary: (2a) forest, (2b) cycle, (2c) 4-petal flower, (2d) complete bipartite.

game compared to the non-attack reachability game. It is thus natural to ask just how dense they can be. In Section 2.4, we prove that (under a mild assumption on the adversary) the equilibria of our game cannot contain more than $2n - 4$ edges when $n \geq 4$. So while these networks can be denser than trees, they remain quite sparse, and thus the threat of attack does not result in too much "over-building" or redundancy of connectivity at equilibrium. Our density upper bound is tight, as the generalized complete bipartite graph in Figure 2d has exactly $2n - 4$ edges.

On the other hand, the examples also show that equilibrium networks can be disconnected (even before the attack) and this might raise concerns regarding the welfare compared to the reachability game. In Section 2.5, we show that for several adversarial attacks, all equilibria which contain at least one edge and at least one immunized vertex (and are thus *non-trivial* as they are different than any equilibrium of the reachability game without attack) are connected and have immunization patterns such that even *after* the attack the network remains highly connected. Hence such equilibria in fact enjoy very good welfare.

2.4. Sparsity

We show that despite the existence of equilibria containing cycles, under a very mild restriction on the adversary, *any* (Nash, swapstable or linkstable) equilibrium network of our game has at most $2n - 4$ edges and is thus quite sparse. Moreover, this upper bound is tight as the generalized complete bipartite graph in Figure 2d has exactly $2n - 4$ edges.

We start by defining a natural restriction on the adversary. We then propose an abstract view of the networks in our game and proceed to show that the abstract network is chord-free in equilibria with respect to the restricted adversary. We finally derive the edge density of the original network by connecting its edge density to the density of the abstract network. We start by defining equivalence classes for networks.

Definition 4. *Let $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ be two networks. G_1 and G_2 are* equivalent *if for all vertices $v \in V$, the connected component of v is the same in both G_1 and G_2 for every possible choice of initial attack vertex in V.*

Based on equivalence, we make the following natural restriction on the adversary.

Assumption 1. *An adversary is* well-behaved *if on any pair of equivalent networks $G_1 = (V, E_1)$ and $G_2 = (V, E_1)$, the probability that a vertex $v \in V$ is chosen for attack, is equal.*

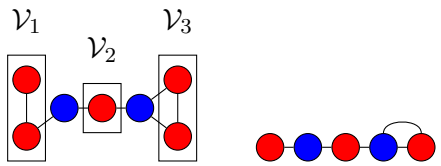
Note that the adversaries in Definitions 1-3 are all well-behaved. We proceed to abstract the network formed by the agents and argue about the edge density in this abstraction.

Let $G = (V, E)$ be any network, $\mathcal{I} \subseteq V$ the immunized vertices in G and $\mathcal{V}_1, \ldots, \mathcal{V}_k$ the vulnerable regions in G. In the abstract network every vulnerable region in G is contracted to a single vertex. More formally, let $G' = (V', E')$ be the abstract network. Define $V' = \mathcal{I} \cup \{u_1, \ldots u_k\}$ where each u_i represents a contracted vulnerable region of G. Moreover, E' is constructed from E as follows. For any edge $(v_1, v_2) \in E$ such that $v_1, v_2 \in \mathcal{I}$ there is an edge $(v_1, v_2) \in E'$. For any edge $(v_1, v_2) \in E$ such that $v_1 \in \mathcal{V}_i$ for some i and $v_2 \in \mathcal{I}$ there is an edge $(u_i, v_2) \in E'$ where u_i denotes the contracted vulnerable region of G that v_1 belongs to. For any edge (v_1, v_2) such that $v_1, v_2 \in \mathcal{V}_i$ for some i there is no edge in G'. We illustrate an example of the original network and the abstract network in Figure 3.

We next show that if G is an equilibrium network then G' is a chord-free graph.

Lemma 1. *Let $G = (V, E)$ be a Nash, swapstable or linkstable equilibrium network and $G' = (V', E')$ the abstraction of G. Then G' is chord-free if the adversary is well-behaved.*

As the next step we bound the edge density of chord-free networks in Theorem 2 using



(a) original network (b) abstract network

Figure 3: Blue: immunized vertices in both networks. Red: vulnerable vertices and regions in the original and abstract networks, respectively.

Theorem 1 from the graph theory literature.

Theorem 1 (Mader [74])**.** *Let $G = (V, E)$ be an undirected graph with minimum degree d. Then there is an edge $(u, v) \in E$ such that there are d vertex-disjoint paths from u to v.*

Theorem 2. *Let $G = (V, E)$ be a chord-free graph on $n \geq 4$ vertices. Then $|E| \leq 2n - 4$.*

Theorem 2 implies the edge density of the abstract network $G' = (V', E')$ is at most $2|V'|-4$. To derive the edge density of the original network, we first show that any vulnerable region in G (contracted vertices in G') is a tree when G is an equilibrium network.

Lemma 2. *Let $G = (V, E)$ be a Nash, swapstable or linkstable equilibrium network. Then all the vulnerable regions in G are trees if the adversary is well-behaved.*

We use Lemmas 1, 2 and Theorem 2 to prove a density bound on the equilibrium networks.

Theorem 3. *Let $G = (V, E)$ be a Nash, swapstable or linkstable equilibrium network on $n \geq 4$ vertices. Then $|E| \leq 2n - 4$ for any well-behaved adversary.*

2.5. Connectivity and Social Welfare in Equilibria

The results of Section 2.4 show that despite the potential presence of cycles at equilibrium, there are still sharp limits on collective expenditure on edges in our game. However, they do not directly lower bound the welfare, due to connectivity concerns: if the graph could become highly fragmented after the attack, or is sufficiently fragmented prior to the attack, the reachability benefits to players could be sharply lower than in the attack-free reachability game. In this section we show that when C_{I} and $C_{\mathrm{E}} > 1$ are both constants with respect to n, none of these concerns are realized in any "interesting" equilibrium network.

In the original reachability game [9], the *maximum* welfare achievable in any equilibrium is $n^2 - O(n)$. Here we will show that the welfare achievable in any "non-trivial" equilibrium is $n^2 - O(n^{5/3})$. Obviously with no restrictions on the adversary and the parameters this cannot be true. Just as in the original game, for $C_{\mathrm{E}} > 1$, the empty graph remains an equilibrium in our game with respect to all the natural adversaries in Section 2.2. The empty graph has a social welfare of only $O(n)$ (each vertex has an expected payoff of $1 - 1/n$). We thus assume the equilibrium network contains at least *one* edge and at least *one* immunized vertex. We refer to all equilibrium networks that satisfy the above assumption as *non-trivial* equilibria. They capture the equilibria that are new to our game compared to the original attack-free setting — the network is not empty, and at least one player has chosen immunization.

Limiting attention to non-trivial equilibria is *necessary* if we hope to guarantee that the welfare at equilibrium is $\Omega(n^2)$ when $C_{\mathrm{E}} > 1$. Without the edge assumption, the empty graph is an equilibrium with respect to several natural adversaries. Furthermore, without the immunization assumption, $n/3$ disjoint components where each component consists of 3 vulnerable vertices is an equilibrium (for carefully chosen C_{E} and C_{I}) with respect to e.g. the maximum carnage adversary. In both cases, the social welfare is only $O(n)$.

Similar to Section 2.4, to get any meaningful results for the welfare we need to restrict the adversary's power. To simplify presentation, for the most of this section we state and analyze our results for the maximum carnage adversary. At the end of this section, we show how these results (or their slight modifications) can be extended to several other adversaries.

Consider any connected component that contains an immunized vertex and an edge in a non-trivial equilibrium network with respect to the maximum carnage adversary. We first show that any targeted region in such component (if exists) has size one when $C_{\mathrm{E}} > 1$.

Lemma 3. *Let G be a non-trivial Nash or swapstable equilibrium network with respect to the maximum carnage adversary. Then in any component of G with at least one immunized vertex and at least one edge, the targeted regions (if they exist) are singletons when $C_E > 1$.*

We then show that non-trivial equilibrium networks with respect to the maximum carnage adversary are connected when $C_{\mathrm{E}} > 1$.

Theorem 4. *Let G be a non-trivial Nash, swapstable or linkstable equilibrium network with respect to the maximum carnage adversary. Then, G is a connected graph when $C_E > 1$.*

Together, Lemma 3 and Theorem 4 imply that any non-trivial equilibrium network with respect to maximum carnage adversary is a connected network with targeted regions of size 1. Finally, we state our main result regarding the welfare in such non-trivial equilibria.

Theorem 5. *Let G be a non-trivial Nash or swapstable equilibrium network on n vertices with respect to the maximum carnage adversary. If C_E and C_I are constants (independent of n) and $C_E > 1$ then the welfare of G is $n^2 - O(n^{5/3})$.*

Lastly, although non-trivial linkstable equilibrium networks with respect to the maximum carnage adversary are connected when $C_{\mathrm{E}} > 1$, the size of targeted regions in such networks can be bigger than 1. So Theorem 5 might not extend to such networks.

Remarks We proved our sparsity result with a rather mild restriction on the adversary. However, we presented our welfare results with respect to a very specific adversary – the maximum carnage adversary. The reader might have noticed that our proofs in this section essentially relied only on the following two properties of the maximum carnage adversary: (1) Adding an edge between any 2 vertices (at least 1 of which is immunized) does not change the distribution of the attack and (2) Breaking a link inside of a targeted region does not increase the probability of attack to the targeted region while at the same time does not decrease the probability of attack to any other vulnerable regions. These same properties hold for the random attack adversary and other adversaries that set the probability of attack to a vulnerable region directly proportional to an increasing function of the size of the vulnerable region. Thus our welfare results extend to random attack adversary and other such adversaries without any modifications.

However, other natural adversaries might not satisfy these properties (e.g. the maximum disruption adversary does not satisfy the first property). While the techniques in the wel-

fare proofs are not directly applicable to such adversaries, it is still possible to reason about the welfare with respect to such adversaries using different techniques e.g. we can show that in any non-trivial and *connected* equilibrium with respect to the maximum disruption adversary, when C_{E} and C_{I} are constants (independent of n) and $C_{\mathrm{E}} > 1$, then the welfare is $n^2 - O(n^{5/3})$. Note that this is slightly weaker than the statement with respect to the maximum carnage adversary, because we cannot show any non-trivial Nash equilibrium network with respect to the maximum disruption adversary is connected when $C_{\mathrm{E}} > 1$. We leave the question of whether arguing about welfare is possible using unified techniques for a wide class of adversaries as future work.

2.6. Simulations

We complement our theory with simulations investigating various properties of swapstable best response dynamics. Again we focused on the maximum carnage adversary and implemented a simulation allowing the specification of the following parameters: number of players n; edge cost C_{E}; immunization cost C_{I}; and initial edge density. The first three of these parameters are as discussed before but the last is new and specific to the simulations. Note that for any $C_{\mathrm{E}} \geq 1$, empty graph is a Nash equilibrium. Thus to sensibly study any type of best response dynamics, it is necessary to "seed" the process with at least some initial connectivity. As for motivation, one could view the initial edge purchases as occurred prior to the introduction of attack and immunization. We examine simulations starting both from very sparse initial connectivity and rather dense initial connectivity, for varying combinations of the other parameters. In all cases the initial connectivity was chosen randomly via the Erdős-Renyi model.

Our simulations proceed in *rounds*, where each round consists of a *swapstable best response update* for all n players in some fixed order. More precisely, in the update for player i we fix the edge and immunization purchases of all other players, and compute the expected payoff of i if she were to alter her current action according to swap deviations stated in Section 2.2. Swapstable dynamics is a rich but "local" best response process, and thus more realistic than

full Nash best response dynamics from a bounded rationality perspective. The phenomena we report on here appear to be qualitatively robust to a variety of natural modifications of the dynamics, such as restriction to linkstable best response, changes to the ordering of updates, and so on. Recall that all of our formal results hold for swapstable as well as Nash equilibria, so the theory remains relevant for the simulations.

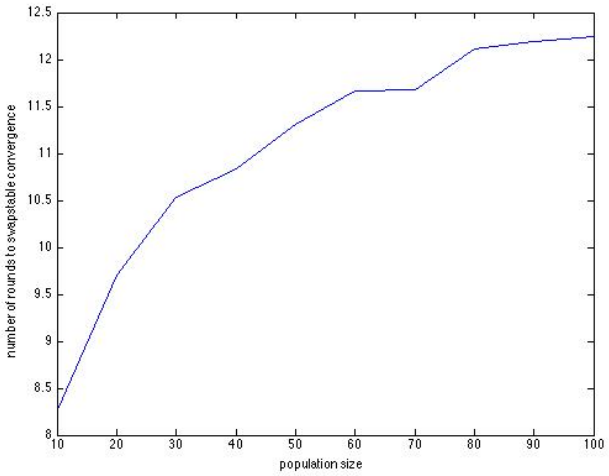


Figure 4: Average number of rounds for swapstable convergence vs. n, for $C_{\mathrm{E}} = C_{\mathrm{I}} = 2$.

The first question that arises in the consideration of any kind of best response dynamic is whether and how quickly it will converge to the corresponding equilibrium notion. Interestingly, empirically it appears that swapstable best response dynamics *always* converges rather rapidly. In Figure 4 we show the average number of rounds to convergence over many trials, starting from dense initial connectivity (average degree 5), for varying values of n. The growth in rounds appears to be strongly sublinear in n (recall that each round updates all n players, so the overall amount of computation is still superlinear in n). Thus we conjecture the general and fast convergence of swapstable dynamics.

In Section 2.3, we gave a number of formal examples of Nash and swapstable equilibria with respect to the maximum carnage adversary. These examples tended to exhibit a large amount of symmetry, especially those containing cycles, due to the large number of cases

that need to be considered in the proofs. Figure 5 shows a sampling of "typical" equilibria found via simulation for $n = 50$ and initial edge density of $1/(2n)$, which exhibit interesting asymmetries and illustrate the effects of the parameters.

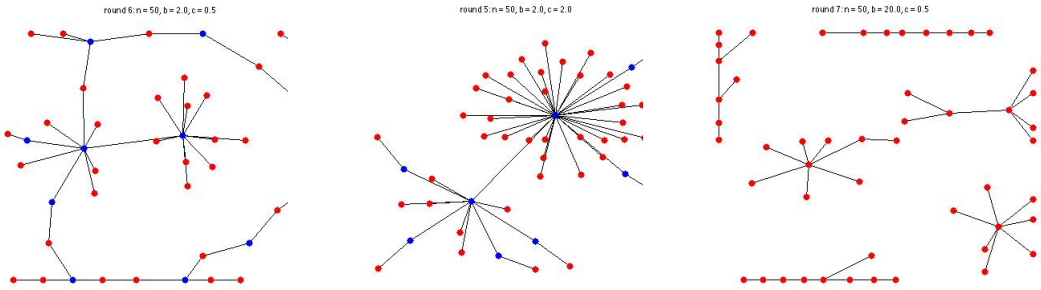


Figure 5: Sample equilibria reached by swapstable best response dynamics for $n = 50$. Left: $C_{\mathrm{E}} = 0.5$, $C_{\mathrm{I}} = 2$. Middle: $C_{\mathrm{E}} = 2$, $C_{\mathrm{I}} = 2$. Right: $C_{\mathrm{E}} = 0.5$, $C_{\mathrm{I}} = 20$.

In the left panel of Figure 5, $C_{\mathrm{E}} = 0.5$ and $C_{\mathrm{I}} = 2$. Thus players have an incentive to buy edges even to isolated vertices as long as they do not increase their vulnerability to the attack. In this regime, despite the initial disconnectedness of the graph, we often see equilibria with a long cycle (as shown), with various tree-like structures attached. In the middle panel we left $C_{\mathrm{I}} = 2$ but increased C_{E} to 2. In this regime cycles are less common due to the higher C_{E}. The equilibrium illustrated is a tree formed by a connected "backbone" of immunized players, each with varying numbers of vulnerable children. Finally, in the right panel we return to inexpensive edges ($C_{\mathrm{E}} = 0.5$), but greatly increased C_{I} to 20. In this regime, we see fragmented equilibria with no immunizations. We note that unlike the right example which is *trivial*, the examples in the left and middle are non-trivial equilibria with high social welfare as predicted by theory.

Figure 5 provides snapshots only at the conclusion of swapstable dynamics while Figure 6 examines entire paths, again at $n = 50$. We started from denser initial graphs (average degree 5), and each panel visualizes a different quantity per number of rounds, for 3 cost regimes: inexpensive cost $C_{\mathrm{E}} = C_{\mathrm{I}} = 2$ (blue); moderate cost $C_{\mathrm{E}} = C_{\mathrm{I}} = 6$ (red); and expensive cost $C_{\mathrm{E}} = C_{\mathrm{I}} = 10$ (green). There are 10 simulations for each cost regime.

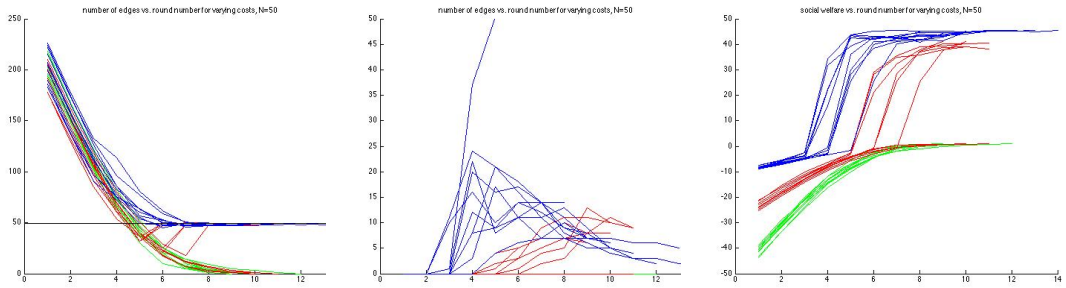


Figure 6: Number of edges (left panel), number of immunizations (middle panel), and average welfare (right panel) vs. number of rounds, for $N = 50$ and varying values for C_{I} and C_{E}. See text for discussion.

In the left panel, we show the evolution of the total number of edges (y axis) in the graph over successive rounds (x axis). In all regimes, there is initially a precipitous decline in connectivity, as the overly dense initial graph cannot be supported at equilibrium. So in the early rounds all players are dropping edges. The ultimate connectivity, however, depends on the cost regime. In the inexpensive regime, connectivity falls monotonically until it levels out very near the threshold for global connectivity at $n - 1$ (horizontal black line), resulting in trees or perhaps just one cycle. In the moderate regime, we see a bifurcation; in some trials, connectivity fall all the way to the empty graph at equilibrium, while in others fall well below the $n - 1$ tree threshold, but then "recover" back to that threshold (which we discuss shortly). In the expensive regime, all trials again result in a monotonic fall of connectivity all the way to the empty graph.

For the same cost regimes and trials, the middle panel shows the number of immunizations purchased over successive rounds. In the inexpensive regime, immunizations, sometimes many, are purchased in early rounds. These act as a "safety net" that prevents connectivity from falling below the tree threshold. Typically immunizations grow initially and then decline. In the moderate regime, we see that the explanation for the connectivity bifurcation discussed above can be traced to immunization decisions. In the trials where connectivity is recovered, some players eventually choose to immunize and thus provide the focal points for edge repurchasing. In many trials resulted in the empty graph, immunizations never

occurred (these remain at $y = 0$). In the expensive regime, no trials are visible because immunizations are never purchased.

Finally, the right panel shows the evolution of the average social welfare per player over successive rounds. In the inexpensive regime, welfare increase slowly and modestly from negative values in the initial graph, then increase dramatically as the benefits of immunization are realized. In the moderate regime, we see a bifurcation of welfare corresponding directly to the bifurcation of connectivity. In the expensive regime, all trials converge from below to the minimum (1-1/n) welfare of the empty graph. Again as theory suggested, the relationship between $C_{\mathrm{E}}, C_{\mathrm{I}}$ and n is determining whether convergence is to a non-trivial equilibrium and thus high social welfare, or to a highly fragmented network with no immunizations and low social welfare.

We conclude by noting that for many parameters, the dynamics above result in heavy-tailed degree distributions — a property commonly observed in large-scale social networks that is easy to capture in stochastic generative models (such as preferential attachment), but more rare in strategic network formation. Across 200 simulations for $n = 100$, $C_{\mathrm{E}} = 0.5$ and $C_{\mathrm{I}} = 2$, we computed the ratio of the maximum to the average degree in each equilibrium found. The lowest, average and maximum ratio observed were 6, 15.8, and 41, respectively (so the highest degree is consistently an order of magnitude greater than the average or more). Moreover, in all 200 trials the highest-degree vertex chose immunization, despite the average rate of immunization of 23% across the population. Thus an amplification process seems to be at work, where vertices that immunize early become the recipients of many edge purchases, since they provide other vertices connectivity benefits that are relatively secure against attack without the cost of immunization.

2.7. A Behavioral Experiment

To complement our theory, we conducted a behavioral experiment on our game with 118 participants. The participants were students in an undergraduate survey course on network

science at the University of Pennsylvania. As training, participants were given a detailed document and lecture on the game, with simple examples of payoffs for players on small graphs under various edge purchase and immunization decisions. Participation was a course requirement, and students were instructed that their grade on the assignment would be exactly equal to their payoffs according to the rules of the game.

The payoffs used the maximum carnage adversary, with costs of $C_{\mathrm{E}} = 5$ and $C_{\mathrm{I}} = 20$. With $n = 118$ participants (so a maximum connectivity benefit of 118 points), it felt that these values made edge purchases and immunization significant expenses and thus worth careful deliberation. Second, running swapstable best response simulations using these values generally resulted in non-trivial equilibria with high welfare, whereas raising C_{E} and C_{I} significantly generally resulted in empty or fragmented graphs with low welfare.

In a game of such complexity, with so many participants, it is unreasonable and uninteresting to formulate the experiment as a one-shot simultaneous move game. Rather, some form of communication must be allowed. We chose to conduct the experiment in a courtyard with the single ground rule that *all conversations be quiet and local* i.e. in order to hear what a participant was saying to others, one should have to stand next to them.

Other than the quiet rule, there were no restrictions on the nature of conversations: participants were free to enter agreements, make promises or threats and move freely. However, it was made clear that any agreements or bargains struck would *not* be enforced by the rules of the experiment (thus were non-binding). Each subject was given a handout that required them to indicate which other subjects they chose to purchase edges to (if any), and whether or not they chose to purchase immunization. The handout contained a list of subject names, along with a unique identification number for each subject used to indicate edge purchases. Thus subjects knew the names of the others as well as their assigned ID numbers. An entire class session was devoted to the experiment, but subjects were free to (irrevocably) turn in their handout at any time and leave. Subjects committed and exited sequentially, and the entire duration was approximately 30 minutes. During the experiment, subjects tended to

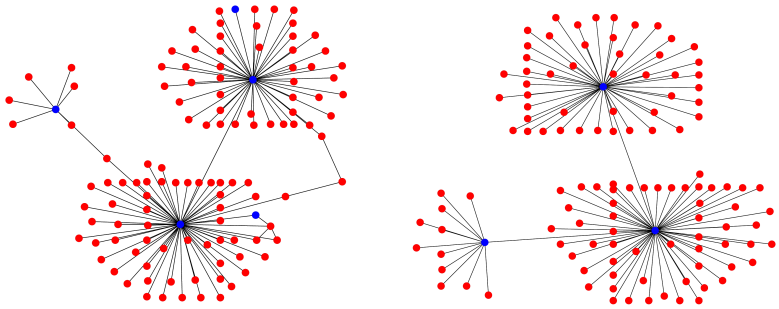


Figure 7: Left: the final undirected network formed by the edge purchases and immunization decisions (blue for immunized, red for vulnerable). Right: a "nearby" Nash network.

gather quickly in small discussion groups that reformed frequently, with subjects moving freely from group to group. It is clear from the outcome that despite adherence to the quiet rule, the subjects engaged in widespread coordination via this rapid mixing.

In the left panel of Figure 7, we show the final undirected network formed by the edge purchases and immunization decisions. The graph is clearly anchored by two main immunized hub vertices, each with many spokes who purchased their single edge to the respective hub. These two large hubs are both directly connected, as well as by a longer "bridge" of three vulnerable vertices. There is also a smaller hub with just a handful of spokes, again connected to one of the larger hubs via a chain of two vulnerable vertices.

For the payoffs, inspection of the network reveals that there are 2 groups of 3 vertices that are the largest vulnerable connected components, and thus are the targets of the attack. These 6 players are each killed with probability 1/2 for a payoff that is only half that of the wealthiest players (the vulnerable spokes of degree 1). In between are the players who purchased immunization including the 3 hubs and 2 immunized spokes. The immunized spoke of the upper hub is unnecessarily so, while the immunized spoke in the lower hub is best responding — had they not purchased immunization, they would have formed a unique largest vulnerable component of size 4 and thus been killed with certainty.

It is striking how many properties the behavioral network shares with the theory: multiple hub-spoke structures with sparse connecting bridges, resulting in high welfare and a heavy-

tailed degree distribution; a couple of cycles. To quantify how far the behavioral network is from equilibrium we use it as the starting point for swapstable best response dynamics and run it until convergence. In the right panel of Figure 7, we show the resulting network reached from the behavioral network in only 4 rounds of swapstable dynamics, and with only 15 of 118 vertices updating their choices. The dynamics *clean up* suboptimal behavioral decisions e.g. the vulnerable bridges between hubs are replaced by direct edges, the other targeted group of 3 spokes drops their fatal edges, and immunizing spokes no longer do so.

Participants were required to complete a survey after the experiment: they were asked to comment on any strategies they contemplated prior to the experiment; whether and how those strategies changed during the experiment; and what strategies or behaviors they observed in other participants.

Many subjects reported entering the experiment with not just a strategy for themselves, but also a "master plan" they hoped to convince others to join. One frequently reported plan involved variations on cycles. Though little thought seems to have been given to coordinating a global ordering in a cycle via only the quiet rule. Another frequently cited plan involved the hub-spoke. Although most strategies are based on abstractions, others reported planning to use social relationships e.g. connecting to students they know.

Of course, of particular interest are the surveys of the hubs. One seems to report an altruistic motivation for purchasing immunization, hoping to maximize welfare. In contrast, the other displays a Machiavellian attitude, willing to immunize in the hopes of creating 3 distinct groups of participants: the "winners" who would connect to the hub; the hub with slightly lower payoff; a large group of "losers" deliberately left out of the hub-spoke structure.

It is clear from the surveys that the word quickly spread during the experiment to connect to hubs and many participants joined though not without some mistrust and hesitation.

We mention some areas for further study. Within our model, the question of whether swapstable best response provably converges (as seen empirically) is open. The benefit function considered here is one of many possible natural choices. It would be interesting to consider other functions. Another extension includes *imperfect* immunization which fails with some probability e.g. as a function of the amount of investment.

We mention two natural variants. The first is the combination of our original model with a standard diffusion model for the spread of attack. For example, combining with the independent cascade model [61], when a targeted vertex is attacked, the infection spreads with probability p along the edges from the attacked point for which both endpoints are unimmunized. This spread then continues until we reach immunized vertices which again act as firewalls. Again different adversaries can have different objectives e.g. the maximum carnage adversary will pick an attack point which maximizes the expected spread. Wang et al. [87] showed that computing the spread in the independent cascade model is #P-complete. This suggests that even before considering the complexity of analysis, agents' reasoning about the choice of attack by the adversary can become quite complicated. Furthermore, due to the probabilistic nature of the spread, it is nontrivial to establish any sparsity properties of the equilibria, because additional overbuilding might occur to hedge against uncertainty of how the infection will spread. Welfare is yet more difficult to analyze; unlike the deterministic spread, it is no longer obvious that a vertex likely to end up in a small component post-attack has a *single* fixed edge purchase that would greatly improve her utility, since different spread patterns can disconnect her from different regions. Regardless of these hurdles, in Chapter 3, we study the variant of the probabilistic spread under the random attack adversary.

The second variant is identical to our current model except that it requires edge purchases to be *bilateral*. In this variant, the concept of equilibrium might be replaced by the notion

of *pairwise stability* (see e.g. [49]). As a majority of our results hinge upon the analysis of unilateral deviations, our current analysis cannot be easily modified to accommodate this change. As a first step towards this goal, the game we study could be modified by adding a *blocking* action with 0 cost while maintaining the unilateral edge formation. Namely for any edge purchased from player i to j, player j can block the edge with no cost. The blocking action removes both the potential connectivity benefit or risk of contagion from edge (i, j) for *both* i and j. The first observation is that there are equilibrium networks in our game which are not equilibria in this new game with blocking. Moreover, we can show that in any equilibrium of the new game, no player blocks any of edges purchased to her. Finally we can show that all the properties of our game (sparsity, connectivity and social welfare) hold in the new game with blocking as well.

Chapter 3

Network Formation Games with Attack and Stochastic Spread

3.1. Introduction

We study a network formation game where strategic agents (vertices on a graph) receive both benefits and costs from forming connections to other agents. While various benefit functions exist in the literature [9, 34], we focus on the *reachability network benefit*. Here, the benefit of an agent is the size of her connected component in the collectively formed graph. This models settings where reachability (rather than centrality) motivates joining the network, e.g. when transmitting packets over technological networks such as the Internet.

Most previous works feature a direct edge cost $C_{\mathrm{E}} > 0$ for forming a link. Goyal et al. [42] (Chapter 2) depart from this notion by studying a game where forming links introduces an additional *indirect* cost by exposing agents to contagious network shocks. These indirect costs can model scenarios such as virus spread through technological or biological networks.

Our work continues this investigation of direct and indirect connection costs. To model the indirect cost we assume that, after network formation, an adversary attacks a single vertex uniformly at random. The attack then kills the vertex and spreads through the network via the independent cascade model according to parameter p [61]. This random attack and probabilistic spread captures the epidemiological quality of virus spread in both biological and technological networks.

At a high level, our work is most closely related to two previous works. Bala and Goyal [9] study a reachability network game without attacks and show a sharp characterization of equilibrium networks: every tree and the empty network can form in equilibria. Goyal

et al. [42] study a reachability network formation game where an adversary inspects the formed network and then deliberately attacks a single vertex in the network. The attack then spreads deterministically to neighboring vertices according to a known rule, while agents may immunize against the attack for a fixed cost. Our game is most similar to the latter setting under a random adversary and high immunization cost. However, in our setting attacks spread probabilistically (through independent cascades) rather than deterministically. This yields an arguably more realistic model of infection spread but incurs additional complexity: computing the expected connectivity benefit of an agent in a given network is now #P-complete [87].

Goyal et al. [42] show that while more diverse equilibrium networks, including ones with multiple cycles, can emerge in addition to trees and the empty graph, the equilibrium networks with n agents will have at most $2n-4$ edges; less than twice the number of edges that can form in the equilibria of the attack-free game. Furthermore, they show that the social welfare is at least $n^2 - o(n^{5/3})$ in non-trivial equilibrium networks. Asymptotically, this is the maximum welfare possible achieved in any nonempty equilibrium of the attack-free game. In the regime where the cost of immunization is high, the game of Goyal et al. [42] only admits disconnected and fragmented equilibrium networks due to deterministic spread of the attack, and the social welfare of the resulting networks may be as low as $\Theta(n)$.

Our Results and Techniques In our game, computing utilities or even verifying network equilibrium is computationally hard. We circumvent this difficulty by proving structural properties for equilibrium networks. First, we provide an upper bound on the edge density in equilibria and show that any equilibrium network on n vertices has $O\left(n \log n/p\right)$ edges.

For constant p this upper bound is tight up to a logarithmic factor. The possibility of over-building therefore differentiates our game from those of Bala and Goyal [9] and Goyal et al. [42], but the extent of over-building is limited.

To prove the density result, we first show that any equilibrium network with more than $\Omega(n \log(n/p))$ edges contains an induced subgraph with large minimum cut size. We then show that if a network has large minimum cut size, in *every* attack (with high probability), either almost all vertices in the network will die or almost all vertices in the network will survive. As a result, any vertex in the induced subgraph can beneficially deviate by dropping an edge. Together, these observations allows us to prove the claimed edge density bound.

Next, we show that any equilibrium network that is nontrivial (i.e. contains at least one edge) also contains a large connected component. Moreover, as long as the network is not too dense, it achieves a constant approximation to the best welfare possible of the attack-free game. More formally, any non-trivial equilibrium network over n vertices contains a connected component of size at least $n/3$. Furthermore, if the number of edges in the network is $O(n/p)$, then the social welfare is $\Omega(n^2)$.

To prove the welfare result, we first show that any agent in a small connected component can increase her connectivity benefits by purchasing an edge to a larger component without significantly increasing her attack risk. This implies the existence of a large connected component. We then use the large component to argue that when the equilibrium network is sparse, the surviving network post-attack still contains a large connected component. This guarantees large social welfare.

While Goyal et al. [42] show robustness of the structural properties of the original reachability game of Bala and Goyal [9] to a variation with attack, deterministic spread and the option of immunization for players, we show robustness in another variant that involves a cascading attack but disallows immunization. However, on the technical front, the tools that we use to prove these robustness results are very different from the previous games.

Organization We introduce our model and discuss the related work in Section 3.2. In Section 3.3 we present examples of equilibrium networks of our game. Sections 3.4 and 3.5 are devoted to the characterization of the edge density and social welfare. We conclude

with directions for future work in Section 3.6.

3.2. Model

We start by formalizing our model for completeness. We assume the n vertices of a graph (network) correspond to individual players. Each player has the choice to purchase edges to other players at a *fixed* cost of $C_{\mathrm{E}} > 0$ per edge. Throughout we assume that C_{E} is a constant independent of n. Furthermore, we use the term *high probability* to refer to probability at least $1 - o(1/n)$ henceforth.

A (pure) *strategy* $s_i \subseteq [n]$ for player i consists of a subset of players to whom player i purchased an edge. Similar to Chapter 2, we assume that edge purchases are unilateral i.e. players do not need approval to purchase an edge to another player but that the connectivity benefits and risks are bilateral.

Let $\mathbf{s} = (s_1, \ldots, s_n)$ denote the strategy profile for all the players. Fixing $\mathbf{s}$, the set of edges purchased by all the players induces an undirected graph. A *game graph* $GG = (V, E)$ is defined to be the undirected graph induced by the edge purchases of all players.

Fixing a game graph G, the adversary selects a *single* vertex $v \in V$ uniformly at random to start the attack. The attack kills v and then spreads according to the independent cascade model with probability $p \in (0, 1)$ [61]. (Throughout we assume that p is a constant independent of the number of players n. We discuss the regime in which p decreases as the number of players increases in Section 3.4.1.) In the independent cascade model, in the first round, the attack spreads independently killing each of the neighbors of the initially attacked vertex v with probability p. In the next round, the spread continues from all the neighbors of v that were killed in the previous round. The spread stops when no new vertex was killed in the last round or when all the vertices are killed.

The adversary's attack can be alternatively described as follows. Fixing a game graph G, let $G[p]$ denote the *random* graph obtained by retaining each edge of G independently with

probability p. The adversary picks a vertex v uniformly at random to start the attack. The attack kills v and all the vertices in the connected component of $G[p]$ that contains v.

Let $CC_i(v)$ denote the *expected* size of the connected component of player i post-attack to a vertex v and we define $CC_v(v)$ to be 0. Then the expected utility (utility for short) of player i in strategy profile $\mathbf{s}$ denoted by $u_i(\mathbf{s})$ is precisely

$$u_i(\mathbf{s}) = \frac{1}{|V|} \sum_{v \in V} CC_i(v) - |s_i| C_{\mathrm{E}}.$$

The sum of utilities of all the players playing $\mathbf{s}$ is defined to be the *social welfare* of $\mathbf{s}$.

Wang et al. [87] show that computing the exact spread of the attack in the independent cascade model is #P-complete in general. This implies that, given a strategy profile $\mathbf{s}$, computing the expected size of the connected component of all vertices (and hence the expected utility of all vertices) is #P-complete. However, an approximation of these quantities can be obtained by Monte Carlo simulation.

We model each of the n players as strategic agents who deterministically choose which edges to purchase. A strategy profile $\mathbf{s}$ is a *pure strategy Nash equilibrium* if, for any player i, fixing the behavior of the other players to be $\mathbf{s}_{-i}$, the expected utility for i, $u_i(\mathbf{s})$, cannot strictly increase when playing any strategy s_i' over s_i. We focus our attention to pure strategy Nash equilibrium (or equilibrium) in this work. Since computing the expected utilities in our game is #P-complete, even verifying that a strategy profile is an equilibrium is #P-complete. Hence as our main contribution, we prove structural properties for the equilibrium networks regardless of this computational barrier.

3.2.1. Related Work

There are two lines of work closely related to ours. First, Bala and Goyal [9] study the attack-free version of our game. They show that equilibrium networks are either trees or the empty network. Also since there is no attack, the social welfare in nonempty equilibrium

networks is asymptotically $n^2 - o(n^2)$.

Second, Goyal et al. [42] study a network formation game where players in addition to having the option of purchasing edges can also purchase immunization from the attack. Since we do not study the effect of immunization purchases in our game, our game corresponds to the regime of parameters in their game where the cost of immunization is so high that no vertex would purchase immunization in equilibria. Moreover, they study several different adversarial attack models and our attack model coincides with their *random attack adversary*. The main difference between our work and theirs is that they assume the attack spreads deterministically while we assume the attack spreads according to the independent cascade model [61]. In many real world scenarios e.g. the spread of contagious disease over the network of people, the spread is *not deterministic*. Hence our work can be seen as a first attempt to make the model of Goyal et al. [42] closer to real world applications. However, the change in the spread of attack comes with a significant increase in the complexity of the game as even computing the utilities of the players in our game is #P-complete. While Friedrich et al. [38] have shown that best responses for players can be computed in polynomial time under various attack models, the question of whether best response dynamics converges to an equilibrium network is open in the model of Goyal et al. [42].

Similar to Goyal et al. [42] we show that diverse equilibrium networks can form in our game. While they show that all equilibrium networks over $n \geq 4$ players have at most $2n-4$ edges, we show that the number of edges in any equilibrium network is at most $O(n \log n)$ and this bound is tight up to a logarithmic factor. Furthermore, Goyal et al. [42] show that the social welfare is asymptotically $n^2 - o(n^2)$ in non-trivial equilibrium networks. Their definition of non-trivial networks requires the network to have at least one immunized vertex and one edge. In the regime where the cost of immunization is high, the game of Goyal et al. [42] only admits disconnected and fragmented equilibrium networks due to the deterministic spread of the attack. Such networks (even excluding the empty graph) can have social welfare as low as $\Theta(n)$. We show that any low density equilibrium network of our game

enjoys a social welfare of $\Theta(n^2)$ as long as the network contains at least one edge.

Kliemann [64] introduced a network formation game with reachability benefits and an attack on the formed network that destroys exactly one link with no further spread. Their equilibrium networks are sparse and also admit high social welfare as removing an edge can create at most two connected components. Kliemann et al. [65] extend this to allow attacks on vertices while focusing on swapstable equilibria.

Blume et al. [16] introduce a game with bilateral edge formation. They assume both edge and link failures can happen simultaneously but independent of the failures so far in the network. These differences make it hard to directly compare the two models.

Finally, network formation games, with a variety of different connectivity benefit models, have been studied extensively in computer science see e.g. [9, 16, 64]. We refer the reader to the related work section in Chapter 2 for a comprehensive summary of other related work especially on the topic of optimal security choices for networks.

3.3. Examples of Equilibrium Networks

In this section we show that a diverse set of topologies can emerge in the equilibrium of our game. Similar to the models of Bala and Goyal [9] and Goyal et al. [42] the empty graph can form in the equilibrium of our game when $C_{\mathrm{E}} \geq 1$. Moreover, similar to both models, trees can form in equilibria (See the left panel of Figure 8). Finally, while Goyal et al. [42] show that in the regime of their game where the cost of immunization is high (so no vertex would immunize) no connected network can form in equilibria due to the deterministic spread of the attack, we show that connected networks indeed can form in our game (See Figure 8).

We remark that pure strategy equilibria exist in all parameter regimes of our game. When $C_{\mathrm{E}} \geq 1$, the empty network can form in equilibria for all p. When $C_{\mathrm{E}} < 1$ a cycle or two disconnected hub-spoke structure of size $n/2$ can form in equilibria depending on whether p is far or close to 1 ($(1 - \omega(1/n))$ and $(1 - o(1))$, respectively).

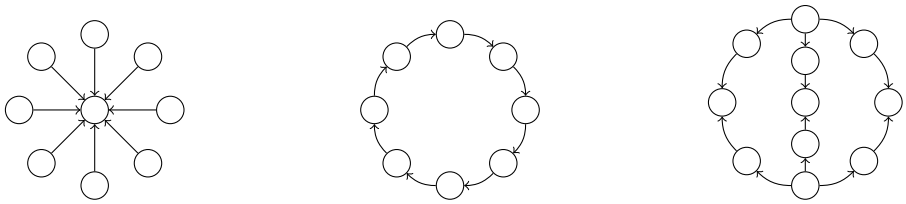


Figure 8: From left to right: hub-spoke, cycle and linear-paths network. A directed arrow determines the vertex that purchases the edge.

Examples in Figure 8 show that denser networks can form in equilibria compared to the model of Bala and Goyal [9] and the high immunization cost regime of the model of Goyal et al. [42]. So we ask how dense equilibrium networks can be in Section 3.4 and show an upper bound of $O(n \log n)$ on the density of the equilibrium networks. Since the examples in Figure 8 have $\Theta(n)$ edges, our upper bound is tight up to a logarithmic factor.

Moreover, while all the equilibrium networks in Figure 8 are connected, there might still exist equilibrium networks in our game that are highly disconnected. In Section 3.5 we show that any equilibrium network with at least one edge contains a large connected component. However, even with the guarantee of a large connected component, there might still be concerns that the equilibrium networks can become highly fragmented after the attack. We show that as long as the equilibrium network is not too dense, the social welfare is lower bounded by $\Theta(n^2)$ i.e. a constant fraction of the social welfare of the attack-free game.

These structural results are obtained despite the computational hardness of equilibrium verification. We view these results as are our most significant technical contributions.

3.4. Edge Density

We now analyze the edge density of equilibrium networks.

Theorem 6. *Any equilibrium network on n vertices has $O(n \log n/p)$ edges.*

The prove Theorem 6, we first show that if G has large enough edge density, then G contains an induced subgraph H whose minimum cut size is large. We then show a large minimum cut size implies that $H[p]$ is connected with high probability and in almost all attacks that

infect a vertex in H, all vertices in H will get infected. So a vertex in H would have a beneficial deviation in the form of dropping an edge; which contradicts the assumption that G was an equilibrium network. This proves that equilibrium networks cannot be too dense.

More formally, we first show in Lemma 4 that if G is *dense enough* it contains a subgraph H with a minimum cut size, denoted by $\alpha(H)$, of at least $\Omega(\log n/p)$.

Lemma 4. *Let $G = (V, E)$ be a graph on n vertices. There exists a constant k such that if $|E| \geq kn \log n/p$ then G contains an induced subgraph H with $\alpha(H) \geq k \log n/p$.*

We then show that if $\alpha(G)$ is $\Omega(\log n/p)$ then with high probability $G[p]$ is connected.

Lemma 5 (Alon [3])**.** *Let $G = (V, E)$ be a graph on n vertices. Then for any constant $b > 0$ there exists a constant $k(b)$ such that if $\alpha(G) \geq k(b) \log n/p$ then with probability at least $1 - n^{-b}$, $G[p]$ is connected.*

We now define a property which we call *almost certain infection* and show that no equilibrium network can contain an induced subgraph satisfying this property.

Definition 5. *Let $G = (V, E)$ be a graph on n vertices and H a subgraph of G on more than one vertex. H has the* almost certain infection *property if whenever any vertex in H is attacked, then with probability at least $1 - o(1/n)$ the attack spreads to every vertex in H.*

Lemma 6. *Let $G = (V, E)$ be an equilibrium network on n vertices. G cannot contain an induced subgraph $H = (V', E')$ such that H satisfies the almost infection property.*

The most interesting regime for the probability of spread p is when p is a constant independent of n. While the upper bound in Theorem 6 holds for all p, it becomes vacuous as p gets small i.e. it becomes bigger than the trivial bound of $n^2/2$ when $p \leq k \log n/n$ for constant k. In Section 3.4.1 we analyze the edge density of equilibrium networks in the regime where $p < 1/n$. We show that the number of edges in any equilibrium network is bounded by $O(n)$ in this regime. The proof utilizes properties of the Galton-Watson branching process and random graph model of Erdös-Rényi, as well as tools from extremal graph theory.

3.4.1. Small p Regime

We now focus on the regime where $p < 1/n$ and prove the following density upper bound.

Theorem 7. *Let $p = \kappa/n$ for some constant $\kappa < 1$. Let $G = (V, E)$ be an equilibrium network over n vertices. Then for sufficiently large n, $|E| \leq \max\{1/C_E, 24000\}n$.*

To prove Theorem 7, we show that if the equilibrium graph has more than $\max\{1/C_{\mathrm{E}}, 24000\}n$ edges, there exists a beneficial deviation in the form of dropping an edge for one of the players. To do so, we need structural results stated in Lemmas 7 and 8. First, consider an edge (u, v) purchased by vertex u. Purchasing this edge would not have increased the connectivity benefit of u unless, after some attack, the edge (u, v) is the only path connecting u to v (and possibly other vertices that are only reachable through v). In Lemma 7 we show that if a graph is dense enough, then there exists an edge (u, v) such that many vertices should be deleted in order to make (u, v) the only remaining path connecting u and v.

Lemma 7. *Let $G = (V, E)$ be a graph on $n > 3\gamma$ vertices with $|E| \geq 2.5\gamma(n - \gamma) - 1$ for some $\gamma \in \mathbb{N}$. Then there exist vertices $v_1, v_2 \in V$ such that $(v_1, v_2) \in E$, and at least $\gamma + 1$ vertices need to be deleted so that the only path from v_1 to v_2 is through the edge (v_1, v_2).*

Second, as described in Section 3.2, the number of vertices that are killed in any attack is the size of the connected component in $G[p]$ that contains the initially attacked vertex. Lemma 8 bounds the size of a randomly chosen connected component in $G[p]$.

Lemma 8. *Let $G = (V, E)$ be an equilibrium network over n vertices with $|E| = kn$ and $\max\{1/C_E, 24000\} \leq k = O(\log n)$. When $p < 1/n$ and n is sufficiently large, the size of the connected component of a randomly chosen vertex v in $G[p]$ is at most $k/3$ with probability at least $1 - 2C_E/(3n)$.*

3.5. Social Welfare

In this section we provide a lower bound on the social welfare of equilibrium networks. Similar to other reachability games, the empty graph can form in equilibrium [9, 42]. Hence

without any further assumptions, no meaningful guarantee on the social welfare can be made. Hence, we focus on non-trivial equilibrium networks defined as follows.

Definition 6. *An equilibrium network is non-trivial if it contains at least one edge.*

Definition 6 rules out the empty network but it is still possible that a non-trivial equilibrium network contains many small connected components or becomes highly fragmented after the attack. In this section we show that none of these concerns materialize. In particular, in Theorem 8, we first show that any non-trivial equilibrium network contains at least one large connected component. We then show in Lemma 9 that when the network is not too dense, the equilibrium network cannot become highly fragmented after the attack. These two observations allows us to prove our social welfare lower bound as stated in Theorem 9.

First we show any non-trivial equilibrium network contains a large connected component.

Theorem 8. *Let $G = (V, E)$ be a non-trivial equilibrium network over n vertices. Then, for sufficiently large n, the largest connected component of G has at least $n/3$ vertices.*

We next present Lemma 9 that describes the relationship between the expected size of the largest connected component of $G[p]$ and the connectivity benefits of the vertices in G.

Lemma 9. *Let $G = (V, E)$ be an equilibrium network over n vertices. Let C be any connected component in G of size n_C. If the expected size of the largest component of $C[p]$ is at most $n_C(1-\epsilon)$, then the expected sum of the connectivity benefits of the vertices in C is at least $(n - n_C)n_C^2/n + \epsilon n_C^3/(3n)$.*

Theorem 8 and Lemma 9 allow us to prove a lower bound on the social welfare.

Theorem 9. *Let $G = (V, E)$ be a non-trivial equilibrium network over n vertices. For any $\epsilon \in (0, 1/8)$ and sufficiently large n, if $|E| < (1-2\epsilon)n\log(1/\epsilon)/p$, then the social welfare of G is at least $\epsilon n^2/3 - O(n/p)$.*

Finally, we remark that unlike the models of Bala and Goyal [9] and Goyal et al. [42], achieving a social welfare of $n^2 - o(n^2)$ is impossible in our game even when restrciting to sparse and non-trivial equilibrium networks. This is formalized in Proposition 1.

Proposition 1. *There exists a non-trivial equilibrium network $G = (V, E)$ over n vertices with $O(n)$ edges such that the social welfare of G is kn^2 for $k < 1$.*

3.6. Conclusions

We studied a natural network formation game where each network connection has the potential to both bring additional utility to an agent as well add to her risk of being infected by a cascading infection attack. We showed that the equilibria resulting from these competing concerns are essentially sparse and containing at most $O(n \log n)$ edges. We also showed that any non-trivial equilibrium network in our game achieves the highest possible social welfare of $\Theta(n^2)$ whenever the equilibrium network has only $O(n)$ edges.

The Price of Anarchy in our model is $\Theta(n)$. To illustrate, consider the $C_{\mathrm{E}} \geq 1$ regime. A central planner can built a cycle or two disconnected hub-spoke structures of size $n/2$ depending on whether the probability of spread of the attack p is low or high, respectively (and both of these structures can also form in equilibrium). Such networks have social welfare of $\Theta(n^2)$. However, the empty network is an equilibrium network in this regime implying a Price of Anarchy of at least $\Theta(n)$ – the worst Price of Anarchy possible. The Price of Stability in our model is $\Theta(1)$ since the social welfare is trivially bounded by n^2 and either of the two equilibrium networks above achieve a social welfare of $\Theta(n^2)$.

Our results suggest several natural questions for future work. Our upper bound of $O(n \log n)$ is a logarithmic factor higher than the densest equilibrium network that we can create. Narrowing this gap is the most interesting open question. Improving our network density upper bound to $O(n)$ edges would immediately imply that all non-trivial equilibrium networks achieve $\Omega(n^2)$ social welfare. Another direction for future work is to analyze how network density and social welfare evolves when agents additionally have an option to invest in immunization that protects them from infections.

Chapter 4

Fairness in Reinforcement Learning

4.1. Introduction

The growing use of machine learning for automated decision-making has raised concerns about the potential for unfairness in learning algorithms and models. In settings as diverse as policing [82], hiring [76], lending [18], and criminal sentencing [10], mounting empirical evidence suggests these concerns are not merely hypothetical [6, 85].

We initiate the study of fairness in reinforcement learning, where an algorithm's choices may influence the state of the world and future rewards. In contrast, previous work on fair machine learning has focused on myopic settings where such influence is absent, e.g. in i.i.d. or no-regret models [30, 35, 45, 52]. The resulting fairness definitions therefore do not generalize well to a reinforcement learning setting, as they do not reason about the effects of short-term actions on long-term rewards. This is relevant for the settings where historical context can have a distinct influence on the future. For concreteness, we consider the specific example of hiring (though other settings such as college admission or lending decisions can be embedded into this framework). Consider a firm aiming to hire employees for a number of positions. The firm might consider a variety of hiring practices, ranging from targeting and hiring applicants from well-understood parts of the applicant pool (which might be a reasonable policy for short-term productivity of its workforce), to exploring a broader class of applicants whose backgrounds might differ from the current set of employees at the company (which might incur short-term productivity and learning costs but eventually lead to a richer and stronger overall applicant pool).

We focus on the standard model of reinforcement learning, in which an algorithm seeks to maximize its discounted sum of rewards in a Markovian decision process (MDP). Through-

out, the reader should interpret the *actions* available to a learning algorithm as corresponding to choices or policies affecting individuals (e.g. which applicants to target and hire). The *reward* for each action should be viewed as the short-term payoff of making the corresponding decision (e.g. the short-term influence on the firm's productivity after hiring any particular candidate). The actions taken by the algorithm affect the underlying state of the system (e.g. the company's demographics as well as the available applicant pool) and therefore in turn will affect the actions and rewards available to the algorithm in the future.

Informally, our fairness notion requires that (with high probability) in state s, an algorithm never chooses an available action a with probability higher than another action a' unless $Q^*(s, a) > Q^*(s, a')$, i.e. the long-term reward of a is greater than that of a'. This definition, adapted from Joseph et al. [52], is *weakly meritocratic*: facing some set of actions, an algorithm must pick a distribution over actions with (weakly) heavier weight on the better actions (in terms of their discounted long-term reward). Correspondingly, a hiring process satisfying our fairness definition cannot probabilistically target one population over another if hiring from either population will have similar long-term benefit to the firm's productivity.

Unfortunately, our first result shows an exponential separation in expected performance between the best unfair algorithm and any algorithm satisfying fairness. This motivates our study of a natural relaxation of (exact) fairness, for which we provide a polynomial time learning algorithm, thus establishing an exponential separation between exact and approximately fair learning in MDPs.

Our Results Throughout, we use *(exact)* fairness to refer to the adaptation of Joseph et al. [52]'s definition defining an action's quality as its potential long-term discounted reward. We also consider two natural relaxations. The first, *approximate-choice fairness*, requires that an algorithm never chooses a worse action with *probability* substantially higher than better actions. The second, *approximate-action fairness*, requires that an algorithm never favors an action of substantially lower *quality* than that of a better action.

The contributions of this paper can be divided into two parts. First, in Section 4.3, we give a lower bound on the time required for a learning algorithm to achieve near-optimality subject to (exact) fairness or approximate-choice fairness. We show that for any constant ϵ, to achieve ϵ-optimality, *(i)* any fair or approximate-choice fair algorithm takes a number of rounds exponential in the number of MDP states and *(ii)* any approximate-action fair algorithm takes a number of rounds exponential in $1/(1-\gamma)$, for discount factor γ.

Second, we present an approximate-action fair algorithm (**Fair-E**^{3}) in Section 4.4 and prove a polynomial upper bound on the time it requires to achieve near-optimality. In particular, we show that for any constant ϵ and any MDP satisfying standard assumptions, **Fair-E**^{3} is an approximate-action fair algorithm achieving ϵ-optimality in a number of rounds that is (necessarily) exponential in $1/(1-\gamma)$ and polynomial in other parameters.

The exponential dependence of **Fair-E**^{3} on $1/(1-\gamma)$ is tight: it matches our lower bound on the time complexity of any approximate-action fair algorithm. Furthermore, our results establish rigorous trade-offs between fairness and performance in reinforcement learning.

4.1.1. Related Work

The most relevant parts of the large body of literature on reinforcement learning focus on constructing learning algorithms with provable performance guarantees. **E**^{3} [59] was the first learning algorithm with a polynomial learning rate, and subsequent work improved this rate (see e.g. Szita and Szepesvári [86]). The study of *robust* MDPs [70, 75, 77] examines MDPs with high parameter uncertainty but generally uses "optimistic" learning strategies that ignore (and often conflict with) fairness and so do not directly apply to this work.

Our work also belongs to a growing literature studying the problem of fairness in machine learning. Early work in data mining [44, 55, 56, 73, 78, 91] considered the question from a primarily empirical standpoint, often using *statistical parity* as a fairness goal. Dwork et al. [30] explicated several drawbacks of statistical parity and instead proposed one of the first broad definitions of algorithmic fairness, formalizing the idea that "similar individuals

should be treated similarly". Recent papers have proven several impossibility results for satisfying different fairness requirements simultaneously [37, 62]. More recently, Hardt et al. [45] proposed new notions of fairness and showed how to achieve these notions via post-processing of a black-box classifier. Woodworth et al. [88] and Zafar et al. [90] further studied these notion theoretically and empirically.

4.1.2. Strengths and Limitations of Our Models

In recognition of the duration and consequence of choices made by a learning algorithm during its learning process – e.g. job applicants not hired – our work departs from previous work and aims to guarantee the fairness of *the learning process itself*. To this end, we adapt the fairness definition of Joseph et al. [52], who studied fairness in the bandit framework and defined fairness with respect to one-step rewards. To capture the desired interaction and evolution of the reinforcement learning setting, we modify this myopic definition and define fairness with respect to long-term rewards: a fair learning algorithm may only choose action a over action a' if a has true long-term reward at least as high as a'. Our contributions thus depart from previous work in reinforcement learning by incorporating a fairness requirement (ruling out existing algorithms which commonly make heavy use of "optimistic" strategies that violates fairness) and depart from previous work in fair learning by requiring "online" fairness in a previously unconsidered reinforcement learning context.

First note that our definition is *weakly meritocratic*: an algorithm satisfying our fairness definition can *never* probabilistically favor a worse option but is not *required* to favor a better option. This confers both strengths and limitations. Our fairness notion still permits a type of "conditional discrimination" in which a fair algorithm favors group A over group B by selecting choices from A when they are superior and randomizing between A and B when choices from B are superior. In this sense, our fairness requirement is relatively minimal, encoding a necessary variant of fairness rather than a sufficient one. This makes our lower bounds and impossibility results (Section 4.3) relatively stronger and upper bounds (Section 4.4) relatively weaker.

Next, our fairness requirement holds (with high probability) across *all* decisions that a fair algorithm makes. We view this strong constraint as worthy of serious consideration, since "forgiving" unfairness during the learning may badly mistreat the training population, especially if the learning process is lengthy or even continual. Additionally, it is unclear how to relax this requirement, even for a small fraction of the algorithm's decisions, without enabling discrimination against a correspondingly small population.

Instead, aiming to preserve the "minimal" spirit of our definition, we consider a relaxation that only prevents an algorithm from favoring a *significantly* worse option over a better option (Section 4.2.1). Hence, approximate-action fairness should be viewed as a weaker constraint: rather than safeguarding against every violation of "fairness", it instead restricts how egregious these violations can be. See Section 4.5 for further relaxations.

4.2. Preliminaries

In this paper we study reinforcement learning in Markov Decision Processes (MDPs). An MDP is a tuple $M = (\mathcal{S}_M, \mathcal{A}_M, P_M, R_M, T, \gamma)$ where $\mathcal{S}_M$ is a set of n *states*, $\mathcal{A}_M$ is a set of k *actions*, T is a *horizon* of a (possibly infinite) number of rounds of activity in M, and γ is a *discount factor*. $P_M : \mathcal{S}_M \times \mathcal{A}_M \to \mathcal{S}_M$ and $R_M : \mathcal{S}_M \to [0, 1]$ denote the *transition probability distribution* and *reward distribution*, respectively. We use $\bar{R}_M$ to denote the mean of R_M. A policy π is a mapping from a history h (the sequence of triples (state, action, reward) observed so far) to a distribution over actions. The discounted state and state-action value functions are denoted by V^π and Q^π, and $V^\pi(s, T)$ represents expected discounted reward of following π from s for T steps. The highest values functions are achieved by the *optimal* policy π^* and are denoted by V^* and Q^* [84]. We use μ^π to denote the stationary distribution of π. Throughout we make the following assumption.

Assumption 2 (Unichain Assumption)**.** *The stationary distribution of any policy in M is independent of its start state.*

We denote the *ϵ-mixing time* of π by T^π_ϵ. Lemma 10 relates the ϵ-mixing time of any policy

π to the number of rounds until the V_M^π values of the visited states by π are close to the expected V_M^π values (under the stationary distribution μ^π).

Lemma 10. *Fix $\epsilon > 0$. For any state s, following π for $T \geq T_\epsilon^\pi$ steps from s satisfies*

$$\mathbb{E}_{s\sim\mu^\pi}\left[V_M^\pi(s)\right] - \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} V_M^\pi(s_t)\right] \leq \frac{\epsilon}{1-\gamma},$$

where s_t is the state visited at time t when following π from s and the expectation in the second term is over the transition function and the randomization of π.

The *horizon time* $H_\epsilon^\gamma := \log\left(\epsilon(1-\gamma)\right)/\log(\gamma)$ of an MDP captures the number of steps an approximately optimal policy must optimize over. The expected discounted reward of any policy after H_ϵ^γ steps approaches the expected asymptotic discounted reward (Kearns and Singh [59], Lemma 2). A learning algorithm $\mathcal{L}$ is a non-stationary policy that at each round takes the entire history and outputs a distribution over actions. We now define a performance measure for learning algorithms.

Definition 7 (ϵ-optimality)**.** *Let $\epsilon > 0$ and $\delta \in (0, 1/2)$. $\mathcal{L}$ achieves* ϵ-optimality *in $\mathcal{T}$ steps if for any $T \geq \mathcal{T}$*

$$\mathbb{E}_{s\sim\mu^*}\left[V_M^*(s)\right] - \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} V_M^*(s_t)\right] \leq \frac{2\epsilon}{1-\gamma}, \tag{4.1}$$

with probability at least $1-\delta$, for s_t the state $\mathcal{L}$ reaches at time t, where the expectation is taken over the transitions and the randomization of $\mathcal{L}$, for any MDP M.

We thus ask that a learning algorithm, after sufficiently many steps, visits states whose values are arbitrarily close to the values of the states visited by the optimal policy. Note that this is stronger than the "hand-raising" notion in Kearns and Singh [59], which only asked that the learning algorithm stop in a state from which discounted return is near-optimal, permitting termination in a state from which the optimal discounted return is poor. In Definition 7, if there are states with poor optimal discounted reward that the optimal policy eventually leaves for better states, so must our algorithms. We also note the following connection between the average V_M^π values of states visited under the stationary

distribution of π (and in particular an optimal policy) and the average undiscounted rewards achieved under the stationary distribution of that policy.

Lemma 11 (Singh [83])**.** *Let $\bar{\boldsymbol{R}}_M$ be the vector of mean rewards in states of M and $\boldsymbol{V}_M^\pi$ the vector of discounted rewards in states under π. Then $\mu^\pi \cdot \bar{\boldsymbol{R}}_M = (1-\gamma)\mu^\pi \cdot \boldsymbol{V}_M^\pi$.*

We design an algorithm which quickly achieves ϵ-optimality and we bound the number of steps $\mathcal{T}$ before this happens by a polynomial in the parameters of M.

4.2.1. Notions of Fairness

We now turn to formal notions of fairness. Translated to our setting, Joseph et al. [52] define action a's quality as the expected immediate reward for choosing a from state s and then require that an algorithm not probabilistically favor a over a' if a has lower expected immediate reward.

However, this naive translation does not adequately capture the structural differences between bandit and MDP settings since present rewards may depend on past choices in MDPs. In particular, defining fairness in terms of immediate rewards would prohibit any policy sacrificing short-term rewards in favor of long-term rewards. This is undesirable, since it is the long-term rewards that matter in reinforcement learning, and optimizing for long-term rewards often necessitates short-term sacrifices. Moreover, the long-term impact of a decision should be considered when arguing about its relative fairness. We will therefore define fairness using the state-action value function Q_M^*.

Definition 8 (Fairness)**.** *$\mathcal{L}$ is* fair *if for all input $\delta > 0$, all M, all rounds t, all states s and all actions a, a'*

$$Q_M^*(s,a) \geq Q_M^*(s,a') \Rightarrow \mathcal{L}(s,a,h_{t-1}) \geq \mathcal{L}(s,a',h_{t-1})$$

with probability at least $1-\delta$ over histories h_{t-1} where $\mathcal{L}(s,a,h)$ denotes the probability $\mathcal{L}$ chooses a from s given history h.

Fairness requires that an algorithm *never* probabilistically favors an action with lower long-term reward over an action with higher long-term reward. In hiring, this means that an algorithm cannot target one applicant population over another unless the targeted population has a higher quality.

In Section 4.3, we show that fairness can be extremely restrictive. Intuitively, $\mathcal{L}$ must play uniformly at random until it has high confidence about the Q^*_M values, in some cases taking exponential time to achieve near-optimality. This motivates relaxing Definition 8. We first relax the *probabilistic* requirement and require only that an algorithm not *substantially* favor a worse action over a better one.

Definition 9 (Approximate-choice Fairness)**.** *$\mathcal{L}$ is* α-choice fair *if for all inputs $\delta > 0$ and $\alpha > 0$: for all M, all rounds t, all states s and actions a, a':*

$$Q^*_M(s,a) \geq Q^*_M(s,a') \Rightarrow \mathcal{L}(s,a,h_{t-1}) \geq \mathcal{L}(s,a',h_{t-1}) - \alpha,$$

with probability of at least $1-\delta$ over histories h_{t-1}. If $\mathcal{L}$ is α-choice fair for any input $\alpha > 0$, we call $\mathcal{L}$ approximate-choice fair*.*

A slight modification of the lower bound for (exact) fairness shows that algorithms satisfying approximate-choice fairness can also require exponential time to achieve near-optimality. We therefore propose an alternative relaxation, where we relax the *quality* requirement. As described in Section 4.1.1, the resulting notion of approximate-action fairness is in some sense the most fitting relaxation of fairness, and is a particularly attractive one because it allows us to give algorithms circumventing the exponential hardness proved for fairness and approximate-choice fairness.

Definition 10 (Approximate-action Fairness)**.** *$\mathcal{L}$ is* α-action fair *if for all inputs $\delta > 0$ and $\alpha > 0$, for all M, all rounds t, all states s and actions a, a':*

$$Q^*_M(s,a) > Q^*_M(s,a') + \alpha \Rightarrow \mathcal{L}(s,a,h_{t-1}) \geq \mathcal{L}(s,a'h_{t-1})$$

with probability of at least $1-\delta$ *over histories* h_{t-1}*. If* $\mathcal{L}$ *is* α*-action fair for any input* $\alpha > 0$*, we call* $\mathcal{L}$ approximate-action fair*.*

Approximate-choice fairness prevents equally good actions from being chosen at very different rates, while approximate-action fairness prevents substantially worse actions from being chosen over better ones. In hiring, an approximately-action fair firm can only (probabilistically) target one population over another if the targeted population is not substantially worse. While this is a weaker guarantee, it at least forces an approximately-action fair algorithm to learn different populations to statistical confidence. This is a step forward from current practices, in which companies have much higher degrees of uncertainty about the quality (and impact) of hiring individuals from under-represented populations. For this reason and the computational benefits mentioned above, our upper bounds will primarily focus on approximate-action fairness.

We now state several useful observations regarding fairness. We note that there always exists a (possibly randomized) optimal policy which is fair; moreover, *any* optimal policy (deterministic or randomized) is approximate-action fair, as is the uniformly random policy.

Finally, we consider a restriction of the actions in an MDP M to nearly-optimal actions (as measured by Q^*_M values).

Definition 11 (Restricted MDP)**.** *The* α-restricted *MDP of* M*, denoted by* M^α*, is identical to* M *except that in each state* s*, the set of available actions are restricted to* $\{a : Q^*_M(s,a) \geq \max_{a' \in \mathcal{A}_M} Q^*_M(s,a') - \alpha \mid a \in \mathcal{A}_M\}$*.*

M^α has the following two properties: (i) any policy in M^α is α-action fair in M and (ii) the optimal policy in M^α is also optimal in M. These properties aid our design of an approximate-action fair algorithm: we construct M^α from estimates of the Q^*_M values (see Section 4.4.3 for more details).

4.3. Lower Bounds

We now demonstrate a stark separation between the performance of learning algorithms with and without fairness. First, we show that neither fair nor approximate-choice fair algorithms achieve near-optimality unless the number of time steps $\mathcal{T}$ is at least $\Omega(k^n)$, exponential in the size of the state space. We then show that any approximate-action fair algorithm requires a number of time steps $\mathcal{T}$ that is at least $\Omega(k^{1/(1-\gamma)})$ to achieve near-optimality. We start by proving a lower bound for fair algorithms.

Theorem 1. *If $\delta < 1/4$, $\gamma > 1/2$ and $\epsilon < 1/8$, no fair algorithm can be ϵ-optimal in $\mathcal{T} = O(k^n)$ steps.*

Standard reinforcement learning algorithms (absent a fairness constraint) learn an ϵ-optimal policy in a number of steps polynomial in n and $1/\epsilon$; Theorem 1 therefore shows a steep cost of imposing fairness. We outline the idea for proof of Theorem 1. For intuition, first consider the special case when the number of actions $k = 2$. We introduce the MDPs witnessing the claim in Theorem 1 for this case.

Definition 12 (Lower Bound Example)**.** *For $\mathcal{A}_M = \{L, R\}$, let $M(x) = (\mathcal{S}_M, \mathcal{A}_M, \mathcal{P}_M, \mathcal{R}_M, T, \gamma, x)$ be an MDP with*

- *for all $i \in [n]$, $P_M(s_i, L, s_1) = P_M(s_i, R, s_j) = 1$ where $j = \min\{i+1, n\}$ and is 0 otherwise.*
- *for $i \in [n-1]$, $R_M(s_i) = 0.5$, and $R_M(s_n) = x$.*

Definition 13. *Let $M(x) = (\mathcal{S}_M, \mathcal{A}_M, P_M, R_M, T, \gamma, x)$ be an MDP where*

- $\mathcal{S}_M = \{s_1, \ldots, s_n\}$.
- $\mathcal{A}_M = \{L, R\}$.
- *for all $i \in [n]$, $P_M(s_i, L, s_1) = P_M(s_i, R, s_j) = 1$ where $j = \min\{i+1, n\}$ and is 0 otherwise (transitions are deterministic).*

- *for* $i \in [n-1]$, $R_M(s_i) = 0.5$, *and* $R_M(s_n) = x$ *(rewards are deterministic).*

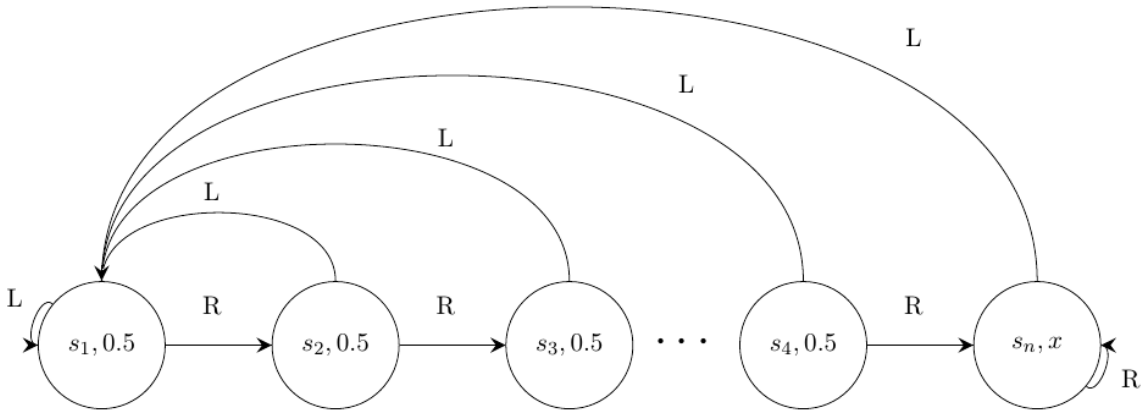


Figure 9: MDP(x): Circles represent states (labels denote the state name and deterministic reward). Arrows represent actions.

Figure 9 illustrates the MDP from Definition 13. All the transitions and rewards in M are deterministic, but the reward at state s_n can be either 1 or $\frac{1}{2}$, and so no algorithm (fair or otherwise) can determine whether the Q^*_M values of all the states are the same or not until it reaches s_n and observes its reward. Until then, fairness requires that the algorithm play all the actions uniformly at random (if the reward at s_n is 1/2, any fair algorithm must play uniformly at random forever). Thus, *any* fair algorithm will take exponential time in the number of states to reach s_n. This can be easily modified for $k > 2$: from each state s_i, $k-1$ of the actions from state s_i (deterministically) return to state s_1 and only one action (deterministically) reaches any other state $s_{\min\{i+1,n\}}$. It will take k^n steps before any fair algorithm reaches s_n and can stop playing uniformly at random (which is necessary for near-optimality). The same example also provides a lower bound of $\Omega((k/(1+k\alpha))^n)$ time steps for approximate-choice fair algorithms as stated in Theorem 2.

Theorem 2. *If* $\delta < 1/4, \alpha < 1/4, \gamma > 1/2$ *and* $\epsilon < 1/8$, *no* α*-choice fair algorithm is* ϵ*-optimal for* $\mathcal{T} = O((k/(1+k\alpha))^n)$ *steps.*

Fairness and approximate-choice fairness are both extremely costly, ruling out polynomial time learning rates. Hence, we focus on approximate-action fairness. Before moving to positive results, we mention that the time complexity of approximate-action fair algorithms

will still suffer from an exponential dependence on $1/(1-\gamma)$.

Theorem 3. *For $\delta < 1/4$, $\alpha < 1/8$, $\gamma > \max(0.9, c)$, $c \in (1/2, 1)$ and $\epsilon < (1 - e^{c-1})/16$, no α-action fair algorithm is ϵ-optimal for $\mathcal{T} = O((k^{1/(1-\gamma)})^c)$ steps.*

The MDP in Figure 9 also witnesses the claim of Theorem 3 when $n = \lceil \log(1/(2\alpha))/(1-\gamma) \rceil$. The discount factor γ is generally taken as a constant, so in most interesting cases $1/(1-\gamma) \ll n$: this lower bound is substantially less stringent than the lower bounds proven for fairness and approximate-choice fairness. Hence, from now on, we focus on designing algorithms satisfying approximate-action fairness with learning rates polynomial in every parameter but $1/(1-\gamma)$, and with tight dependence on $1/(1-\gamma)$.

4.4. A Fair and Efficient Learning Algorithm

We now present an approximate-action fair algorithm, **Fair-E**^{3} with the performance guarantees stated below.

Theorem 4. *Given $\epsilon > 0$, $\alpha > 0$, $\delta \in (0, 1/2)$ and $\gamma \in [0, 1)$ as inputs, **Fair-E**^{3} is an α-action fair algorithm which achieves ϵ-optimality after*

$$\mathcal{T} = \tilde{O}\left(\frac{n^5 T_\epsilon^* k^{\frac{1}{1-\gamma}+5}}{\min\{\alpha^4, \epsilon^4\}\epsilon^2 (1-\gamma)^{12}}\right) \tag{4.2}$$

steps where $\tilde{O}$ hides poly-logarithmic terms.

The running time of **Fair-E**^{3} (which we have not attempted to optimize) is polynomial in all the parameters of the MDP except $1/(1-\gamma)$; Theorem 3 implies that this exponential dependence on $1/(1-\gamma)$ is necessary.

Several more recent algorithms (e.g. R-MAX [17]) have improved upon the performance of **E**^{3}. We adapted **E**^{3} primarily for its simplicity. While the machinery required to properly balance fairness and performance is somewhat involved, the basic ideas of our adaptation are intuitive. We further note that subsequent algorithms improving on **E**^{3} tend to heavily leverage the principle of "optimism in face of uncertainty": such behavior often violates

fairness, which generally requires *uniformity* in the face of uncertainty. Thus, adapting these algorithms to satisfy fairness is more difficult. This in particular suggests $\mathbf{E}^3$ as an apt starting point for designing a fair planning algorithm.

The remainder of this section will explain **Fair-E**^{3}, beginning with a high-level description in Section 4.4.1. We then define the "known" states **Fair-E**^{3} uses to plan in Section 4.4.2, explain this planning process in Section 4.4.3, and bring this all together to prove **Fair-E**^{3}'s fairness and performance guarantees in Section 4.4.4.

*4.4.1. Informal Description of **Fair-E**^{3}*

Fair-E^3 relies on the notion of "known" states. A state s is defined to be *known* after all actions have been chosen from s enough times to confidently estimate relevant reward distributions, transition probabilities, and Q^π_M values for each action. At each time t, **Fair-E**^{3} then uses known states to reason about the MDP as follows:

- If in an unknown state, take a uniformly random trajectory of length H^γ_ϵ.
- If in a known state, compute *(i)* an exploration policy which escapes to an unknown state quickly and p, the probability that this policy reaches an unknown state within $2T^*_\epsilon$ steps, and *(ii)* an exploitation policy which is near-optimal in the known states of M.
  - If p is large enough, follow the exploration policy; otherwise, follow the exploitation policy.

Fair-E^3 thus relies on known states to balance exploration and exploitation in a reliable way. While **Fair-E**^{3} and $\mathbf{E}^3$ share this general idea, fairness forces **Fair-E**^{3} to more delicately balance exploration and exploitation. For example, while both algorithms explore until states become "known", the definition of a known state must be much stronger in **Fair-E**^{3} than in $\mathbf{E}^3$ because **Fair-E**^{3} additionally requires accurate estimates of actions' Q^π_M values in order to make decisions without violating fairness. For this reason, **Fair-E**^{3} replaces the

deterministic exploratory actions of $\mathbf{E}^3$ with random trajectories of actions from unknown states. These random trajectories are then used to estimate the necessary Q^{π}_{M} values.

In a similar vein, **Fair-E**^{3} requires particular care in computing exploration and exploitation policies, and must restrict the set of such policies to fair exploration and fair exploitation policies. Correctly formulating this restriction process to balance fairness and performance relies heavily on the observations about the relationship between fairness and performance provided in Section 4.2.1.

*4.4.2. Known States in **Fair-E**^{3}*

We now formally define the notion of known states for **Fair-E**^{3}. We say a state s becomes known when one can compute good estimates of *(i)* $R_M(s)$ and $P_M(s,a)$ for all a, and *(ii)* $Q^*_M(s,a)$ for all a.

Definition 14 (Known State)**.** *Let*

$$m_1 = O\left(k^{H^{\gamma}_{\epsilon}+3} n \left(\frac{1}{(1-\gamma)\,\alpha}\right)^2 \log\left(\frac{k}{\delta}\right)\right) \text{ and } m_2 = O\left(\left(\frac{n}{\min\{\epsilon,\alpha\}}\right)^4 H^{\gamma 8}_{\epsilon} \log\left(\frac{1}{\delta}\right)\right).$$

A state s becomes known *after taking*

$$m_Q := k \cdot \max\{m_1, m_2\} \tag{4.3}$$

length-H^{γ}_{ϵ} random trajectories from s.

It remains to show that motivating conditions *(i)* and *(ii)* indeed hold for our formal definition of a known state. Informally, m_1 random trajectories suffice to ensure that we have accurate estimates of all $Q^*_M(s,a)$ values, and m_2 random trajectories suffice to ensure accurate estimates of the transition probabilities and rewards.

To formalize condition *(i)*, we rely on Theorem 5, connecting the number of random trajectories taken from s to the accuracy of the empirical V^{π}_{M} estimates.

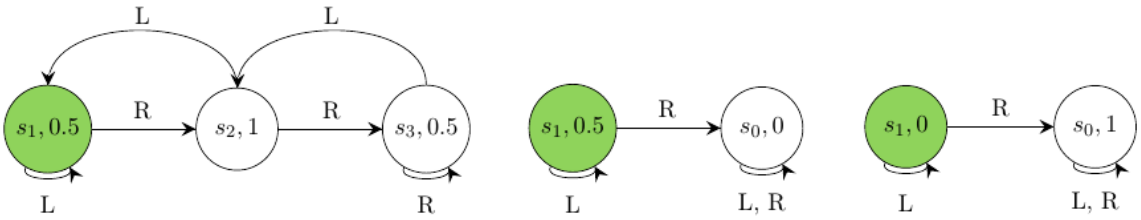


Figure 10: Left: An MDP M with actions L and R and deterministic transition functions and rewards. Green denotes the set of known states Γ. Middle: M_Γ. Right: $M_{[n]\setminus\Gamma}$.

Theorem 5 (Theorem 5.5, Kearns et al. [60]). *For any state s and $\alpha > 0$, after*

$$m = O\left(k^{H_\epsilon^\gamma + 3}\left(\frac{1}{(1-\gamma)\,\alpha}\right)^2 \log\left(\frac{|\Pi|}{\delta}\right)\right)$$

random trajectories of length H_ϵ^γ from s, with probability of at least $1-\delta$, we can compute estimates $\hat{V}_M^\pi$ such that $|V_M^\pi(s) - \hat{V}_M^\pi(s)| \leq \alpha$, simultaneously for all $\pi \in \Pi$.

Theorem 5 enables us to translate between the number of trajectories taken from a state and the uncertainty about its V_M^π values for all policies (including π^* and hence V_M^*). Since $|\Pi| = k^n$, we substitute $\log(|\Pi|) = n\log(k)$. To estimate $Q_M^*(s,a)$ values using the $V_M^*(s)$ values we increase the number of necessary length-H_ϵ^γ random trajectories by a factor of k.

For condition *(ii)*, we adapt the analysis of $\mathbf{E}^3$ [59], which states that if each action in a state s is taken m_2 times, then the transition probabilities and reward in state s can be estimated accurately (see Section 4.4.4).

4.4.3. Planning in ***Fair-E***^{3}

We now formalize the planning steps in **Fair-E**^{3} from known states.

Fair-E^3 constructs two ancillary MDPs for planning: M_Γ is the *exploitation* MDP, in which the unknown states of M are condensed into a single absorbing state s_0 with no reward. In the known states Γ, transitions are kept intact and the rewards are deterministically set to their mean value. M_Γ thus incentivizes exploitation by giving reward only for staying

within known states. In contrast, $M_{[n]\setminus\Gamma}$ is the *exploration* MDP, identical to M_Γ except for the rewards. The rewards in the known states Γ are set to 0 and the reward in s_0 is set to 1. $M_{[n]\setminus\Gamma}$ then incentivizes exploration by giving reward only for escaping to unknown states. See the middle (right) panel of Figure 10 for an illustration of M_Γ ($M_{[n]\setminus\Gamma}$).

Fair-E^3 uses these constructed MDPs to plan according to the following natural idea: when in a known state, **Fair-E^3** constructs $\hat{M}_\Gamma$ and $\hat{M}_{[n]\setminus\Gamma}$ based on the estimated transition and rewards observed so far, and then uses these to compute additional restricted MDPs $\hat{M}^\alpha_\Gamma$ and $\hat{M}^\alpha_{[n]\setminus\Gamma}$ for approximate-action fairness. **Fair-E^3** then uses these restricted MDPs to choose between exploration and exploitation.

More formally, if the optimal policy in $\hat{M}^\alpha_{[n]\setminus\Gamma}$ escapes to the absorbing state of M_Γ with high enough probability within $2T^*_\epsilon$ steps (T^*_ϵ can be guessed by the doubling trick), then **Fair-E^3** explores by following that policy. Otherwise, **Fair-E^3** exploits by following the optimal policy in $\hat{M}^\alpha_\Gamma$ for T^*_ϵ steps. While following either of these policies, whenever **Fair-E^3** encounters an unknown state, it stops following the policy and proceeds by taking a length-H^γ_ϵ random trajectory.

*4.4.4. Analysis of **Fair-E^3***

We next formally analyze **Fair-E^3** and prove Theorem 4. We begin by proving that M^α_Γ is useful in the following sense: M^α_Γ has at least one of an exploitation policy achieving high reward or an exploration policy that quickly reaches an unknown state in M.

Lemma 12 (Exploit or Explore Lemma)**.** *For any state $s \in \Gamma$, $\beta \in (0,1)$ and any $T > 0$ at least one of the statements below holds:*

- *there exists an* exploitation policy π *in M^α_Γ such that*

$$\max_{\bar{\pi}\in\Pi} \mathbb{E}\sum_{t=1}^{T} V^{\bar{\pi}}_M\left(\bar{\pi}^t(s), T\right) - \mathbb{E}\sum_{t=1}^{T} V^{\pi}_{M_\Gamma}\left(\pi^t(s), T\right) \le \beta T$$

  where the random variables $\pi^t(s)$ and $\bar{\pi}^t(s)$ denote the states reached from s after

following π and $\bar{\pi}$ for t steps, respectively.

- *there exists an* exploration policy π *in* M_Γ^α *such that the probability that a walk of* $2T$ *steps from* s *following* π *will terminate in* s_0 *exceeds* β/T.

We can use this fact to reason about exploration as follows. First, since the optimal policy in M is approximate-action fair, if the optimal policy stays in the set of M's known states M_Γ, then following the optimal policy in M_Γ^α is both optimal and approximate-action fair.

However, if instead the optimal policy in M quickly escapes to an unknown state in M, the optimal policy in M_Γ^α may not be able to compete with the optimal policy in M. Ignoring fairness, one natural way of computing an escape policy to "keep up" with the optimal policy is to compute the optimal policy in $M_{[n]\setminus\Gamma}$. Unfortunately, following this escape policy might violate approximate-action fairness – high-quality actions might be ignored in lieu of low-quality exploratory actions that quickly reach the unknown states of M. Instead, we compute an escape policy in $M_{[n]\setminus\Gamma}^\alpha$ and show that if no near-optimal exploitation policy exists in M_Γ, then the optimal policy in $M_{[n]\setminus\Gamma}^\alpha$ (which is fair by construction) quickly escapes to the unknown states of M.

Next, in order for **Fair-E**^{3} to check whether the optimal policy in $M_{[n]\setminus\Gamma}^\alpha$ quickly reaches the absorbing state of M_Γ with significant probability, **Fair-E**^{3} simulates the execution of the optimal policy of $M_{[n]\setminus\Gamma}^\alpha$ for $2T_\epsilon^*$ steps from the known state s in M_Γ^α several times, counting the ratio of the runs ending in s_0, and applying a Chernoff bound.

Having discussed exploration, it remains to show that the exploitation policy described in Lemma 12 satisfies ϵ-optimality as defined in Definition 7. By setting $T \geq T_\epsilon^*$ in Lemma 12 and applying Lemmas 10 and 13, we can prove Corollary 1 regarding this exploitation policy.

Corollary 1. *For any state $s \in \Gamma$ and $T \geq T_\epsilon^*$ if there exists an exploitation policy π in M_Γ^α then*

$$\left| \frac{1}{T} \mathbb{E} \sum_{t=1}^{T} V_M^\pi \left(\pi^t(s), T \right) - \mathbb{E}_{s \sim \mu^*} V_M^*(s) \right| \leq \frac{\epsilon}{1-\gamma}.$$

Finally, we have so far elided the fact that **Fair-E**^{3} only has access to the *empirically estimated* MDPs $\hat{M}^{\alpha}_{\Gamma}$ and $\hat{M}^{\alpha}_{[n]\setminus\Gamma}$. We remedy this issue by showing that the behavior of any policy π in $\hat{M}^{\alpha}_{\Gamma}$ (and $\hat{M}^{\alpha}_{[n]\setminus\Gamma}$) is similar to the behavior of π in M^{α}_{Γ} (and $M^{\alpha}_{[n]\setminus\Gamma}$). To do so, we prove a stronger claim: the behavior of any π in $\hat{M}_{\Gamma}$ (and $\hat{M}_{[n]\setminus\Gamma}$) is similar to the behavior of π in M_{Γ} (and $M_{[n]\setminus\Gamma}$).

Lemma 13. *Let Γ be the set of known states and $\hat{M}_{\Gamma}$ the approximation to M_{Γ}. Then for any state $s \in \Gamma$, any action a and any policy π, with probability at least $1-\delta$:*

1. $V^{\pi}_{M_{\Gamma}}(s) - \min\{\frac{\alpha}{2}, \epsilon\} \leq V^{\pi}_{\hat{M}_{\Gamma}}(s) \leq V^{\pi}_{M_{\Gamma}}(s) + \min\{\frac{\alpha}{2}, \epsilon\}$,

2. $Q^{\pi}_{M_{\Gamma}}(s,a) - \min\{\frac{\alpha}{2}, \epsilon\} \leq Q^{\pi}_{\hat{M}_{\Gamma}}(s,a) \leq Q^{\pi}_{M_{\Gamma}}(s,a) + \min\{\frac{\alpha}{2}, \epsilon\}$.

4.5. Discussion and Future Work

Our work leaves open several interesting questions. For example, we give an algorithm that has an undesirable exponential dependence on $1/(1-\gamma)$, but we show that this dependence is unavoidable for any approximate-action fair algorithm. Without fairness, near-optimality in learning can be achieved in time that is polynomial in *all* of the parameters of the underlying MDP. So, we can ask: does there exist a *meaningful* fairness notion that enables reinforcement learning in time polynomial in all parameters?

Moreover, our fairness definitions remain open to further modulation. It remains unclear whether one can *strengthen* our fairness guarantee to bind across time rather than simply across actions available at the moment without large performance tradeoffs. Similarly, it is not obvious whether one can gain performance by *relaxing* the every-step nature of our fairness guarantee in a way that still forbids discrimination. These and other considerations suggest many questions for further study; we therefore position our work as a first cut for incorporating fairness into a reinforcement learning setting.

Chapter 5

Fairness in Regression

5.1. Introduction

The widespread use of machine learning to make consequential decisions about individual citizens (including in domains such as credit, employment, education and criminal sentencing [10, 18, 76, 82]) has been accompanied by increased reports of instances in which the algorithms and models employed can be unfair or discriminatory in a variety of ways [6, 85]. As a result, research on fairness in machine learning and statistics has seen rapid growth in recent years [1, 19, 20, 24, 30, 35, 36, 44, 45, 53–56, 73, 78], and several mathematical formulations have been proposed as metrics of (un)fairness for a number of different learning frameworks. While much of the attention to date has focused on (binary) classification settings, where standard fairness notions include equal false positive or negative rates across different populations, less attention has been paid to fairness in (linear and logistic) regression settings, where the target and/or predicted values are continuous, and the same value may not occur even twice in the training data.

In this work, we introduce a rich family of fairness metrics for regression models that take the form of a fairness regularizer and apply them to the standard loss functions for linear and logistic regression. Since these loss functions and our fairness regularizer are convex, the combined objective functions obtained from our framework are also convex, and thus permit efficient optimization. Furthermore, our family of fairness metrics covers the spectrum from the type of *group* fairness that is common in classification formulations (where e.g. false arrests in one racial group can be "compensated" for by false arrests in another racial group) to much stronger notions of *individual* fairness (where such cancellations are forbidden, and every injustice is charged to the model). Intermediate fairness notions are also covered. Our framework also permits one to either forbid the use of a "protected" variable (such as

race), by demanding that a single model be learned across all groups, or to build different group-dependent models.

Most importantly, by varying the weight on the fairness regularizer, our framework permits us to compute the entire "Pareto curve" or efficient frontier of the trade-off between predictive accuracy and fairness. Such curves are especially important to examine and understand in a domain-specific manner: since demanding fairness of models will always come at a cost of reduced predictive accuracy [25, 36, 50, 90], it behooves practitioners working with fairness-sensitive data sets to understand just how mild or severe this trade-off is in their particular arena, permitting them to make informed modeling and policy decisions.

Our central results take the form of an extensive comparative empirical case study across six distinct datasets in which fairness is a primary concern. For each of these datasets, we compute and examine the corresponding fairness-accuracy efficiency frontier. We introduce an intuitive quantity called the *Price of Fairness (PoF)*, which numerically quantifies the extent to which increased fairness degrades accuracy. We compare the PoF across datasets, fairness notions, and treatments of protected variables.

Our primary contributions are:

- The introduction of a flexible but convex family of fairness regularizers of varying strength that spans the spectrum from group to individual fairness.
- The introduction of a quantitative, data-dependent measure of the severity of the accuracy-fairness tradeoff.
- An extensive empirical comparative study across six fairness-sensitive data sets.

While our empirical study does reveal some reasonably consistent findings across datasets (e.g. efficiency curves show broadly similar shapes; PoF generally higher for individual fairness than group; somewhat surprisingly, PoF not generally improved much when using protected variables), perhaps the most important message is a cautionary one: the detailed

trade-off between accuracy and fairness, and the comparison of different fairness notions, appears to be quite domain-dependent and lacking prescriptive "universals". This is perhaps consistent with the emerging theoretical literature demonstrating the lack of a single "right" definition of fairness [24, 37, 62], and our work adds evidence to the view that fairness is a topic demanding careful domain-specific considerations.

5.2. The Regression Setting

Consider the (linear and logit) regression setting: denote the *explanatory* variables (or *instances*) by $\mathbf{x} \in \mathcal{X} = \mathbb{R}^d$ and the *target* variables (or *labels*) by $y \in \mathcal{Y} = [-1, 1]$. For both linear and logit models, the target values are continuous. Let $\mathcal{P}$ denote the joint distribution over $\mathcal{X} \times \mathcal{Y}$. Suppose every instance $\mathbf{x}$ belongs to exactly one of 2 groups, denoted by 1 and 2. This partition of $\mathcal{X}$ into groups (e.g. into different races or genders) is encoded in a "sensitive" feature $\mathcal{X}_{d+1}$. Let $S = \{(\mathbf{x}_i, y_i)\}_{i=1}^n\}$ be a training set of n samples drawn i.i.d. from $\mathcal{P}$, separated by groups into S_1 and S_2. Let $n_1 = |S_1|$ and $n_2 = |S_2|$. $(n = n_1 + n_2.)$

This work studies the trade-off between fairness and accuracy for the class of linear and logit regression models. Given a pair of explanatory and target variables $(\mathbf{x}, y)$, we treat y as the ground truth description of $\mathbf{x}$'s merit for the regression task at hand: two pairs $(\mathbf{x}, y), (\mathbf{x}', y')$ with $y \approx y'$ have similar observed outcomes. We aim to design models which treat two such instances with similar observed outcomes similarly, a notion we refer to as *fairness* with respect to the ground truth. For a given accuracy loss ℓ and fairness loss (or *penalty*) f, we define the λ-weighted fairness loss of a regressor $\mathbf{w}$ on a distribution $\mathcal{P}$ to be $\ell_{\mathcal{P}}(\mathbf{w}) + \lambda f_{\mathcal{P}}(\mathbf{w})$. For our sample S, we analogously define the λ-weighted fairness training loss of $\mathbf{w}$ as $\ell(\mathbf{w}, S) + \lambda f(\mathbf{w}, S)$. For linear regression, we let ℓ be mean-squared error; for logistic regression, we let ℓ be the standard log loss. Finally, we use ℓ_2 regularization for both models, so the overall loss is then $\ell_{\mathcal{P}}(\mathbf{w}) + \lambda f_{\mathcal{P}}(\mathbf{w}) + \gamma||\mathbf{w}||_2$.

5.2.1. A Convex Family of Fairness Regularizers

Our formal definitions of fairness all measure how similarly a model treats two similarly labeled instances, one from group 1 and one from group 2. In particular, all of our definitions have a term for each "cross-group" *pair* of instances/labels, weighted as a function of $|y_i - y_j|$ and also by $|\mathbf{w}^\top \mathbf{x}_i - \mathbf{w}^\top \mathbf{x}_j|$. For shorthand, we will refer to pairs of instances (one from each group) as *cross pairs*, and cross pairs with similar labels as similar cross pairs. Each of the below fairness definitions differs in precisely which cross pair disparities can counteract one another. In one extreme (individual fairness in Equation (5.1)), every cross pair disparity increases the fairness penalty of a model. In the other (group fairness in Equation (5.2)), making higher predictions for the group 1 instance of a similar cross pair can be somewhat counterbalanced by making a higher prediction for the group 2 instance of a different similar cross pair. Our notions of fairness for regression align closely to individual and group fairness definitions for classification, both common threads in the fairness literature.

Remark 1. *We assume the sensitive feature $\mathcal{X}_{d+1}$ is available to the learning procedure in one of two ways. In the first setting, which we call the "single model" setting, we assume the algorithm builds a single linear model $\mathbf{w}$ for all of $\mathcal{X}$ (over all but the sensitive features), but can measure the empirical fairness loss of $\mathbf{w}$ using the sensitive feature. In the second setting, which we call the "separate models" setting, we allow the algorithm to build two distinct linear models $\mathbf{w}_1, \mathbf{w}_2$ for the two groups, $\mathbf{w}_g$ based on S_g, thus directly observing the sensitive feature when building these models.*

We specialize our fairness penalties for the single model setting, but we extend them to the separate models setting, by replacing w with w_g when applied to a member of group g.

Individual Fairness The first fairness penalty we propose is the following:

$$f_1(\mathbf{w}, S) = \frac{1}{n_1 n_2} \sum_{\substack{(\mathbf{x}_i, y_i) \in S_1 \\ (\mathbf{x}_j, y_j) \in S_2}} d(y_i, y_j) \big(\mathbf{w}^\top \mathbf{x}_i - \mathbf{w}^\top \mathbf{x}_j\big)^2, \tag{5.1}$$

for some fixed non-negative function d, which we assume is decreasing in $|y_i - y_j|$ (see Section 5.4 for more details). Since $d(y_i, y_j)$ does not depend upon the decision variables ($\mathbf{w}$), one can treat these values as constants in an optimization procedure for selecting $\mathbf{w}$.

The penalty f_1 corresponds to *individual fairness*; for every cross pair $(\mathbf{x}, y) \in S_1, (\mathbf{x}', y') \in S_2$, a model $\mathbf{w}$ is penalized for how differently it treats $\mathbf{x}$ and $\mathbf{x}'$ (weighted by a function of $|y - y'|$). No cancellation occurs: the penalty for overestimating several of one group's labels cannot be mitigated by overestimating several of the other group's labels.

Group Fairness The second fairness penalty we propose is the following:

$$f_2(\mathbf{w}, S) = \Big(\frac{1}{n_1 n_2} \sum_{\substack{(\mathbf{x}_i, y_i) \in S_1 \\ (\mathbf{x}_j, y_j) \in S_2}} d(y_i, y_j)\big(\mathbf{w}^\top \mathbf{x}_i - \mathbf{w}^\top \mathbf{x}_j\big)\Big)^2. \tag{5.2}$$

The penalty f_2 corresponds to *group fairness*: on average, the two groups' instances should have similar labels (weighted by the nearness of the labels of the instances). Unlike f_1, the penalty f_2 allows for *compensation*: informally, if the model over-values some instances of group 1 relative to group 2 in similar cross pairs, it can compensate on other similar cross-pairs by over-valuing those instances from group 2 relative to group 1.

In both of the above formulations, for any cross pair $(\mathbf{x}_i, y_i) \in S_1$ and $(\mathbf{x}_j, y_j) \in S_2$, any regressor $\mathbf{w}$ will have penalty that increases as $\big|\mathbf{w}^\top \mathbf{x}_i - \mathbf{w}^\top \mathbf{x}_j\big|$ increases, weighted by $d(y_i, y_j)$. If the cross pair is similar (y_i is close to y_j and $d(y_i, y_j)$ is large), a regressor which makes very different predictions for $\mathbf{x}_i$ and $\mathbf{x}_j$ will incur large loss. If the cross pair is less similar (y_i is far from y_j and $d(y_i, y_j)$ is smaller), there is less penalty for having a regressor for which $\big|\mathbf{w}^\top \mathbf{x}_i - \mathbf{w}^\top \mathbf{x}_j\big|$ is large.

Hybrid notions of fairness Note that group and individual fairness correspond to two extremes: in one extreme the fairness penalty considers each cross pair separately and in the other one the fairness penalty considers all the cross pairs together. Mathematically one

could define different notions of fairness by grouping the cross pairs in different manners or even restrict the fairness penalty only on a subset of cross pairs (called *bucketing*). In particular, for binary labeled data (where $\mathcal{Y} = \{-1, 1\}$) one natural choice is to group the cross pairs based on their labels. This would result in the following definition of fairness which we call hybrid fairness:

$$f_3(\mathbf{w}, S) = \Big(\sum_{\substack{(\mathbf{x}_i, y_i) \in S_1 \\ (\mathbf{x}_j, y_j) \in S_2 \\ y_i = y_j = 1}} \frac{d(y_i, y_j)\big(\mathbf{w}^\top \mathbf{x}_i - \mathbf{w}^\top \mathbf{x}_j\big)}{n_{1,1} n_{2,1}} \Big)^2 + \Big(\sum_{\substack{(\mathbf{x}_i, y_i) \in S_1 \\ (\mathbf{x}_j, y_j) \in S_2 \\ y_i = y_j = -1}} \frac{d(y_i, y_j)\big(\mathbf{w}^\top \mathbf{x}_i - \mathbf{w}^\top \mathbf{x}_j\big)}{n_{1,-1} n_{2,-1}} \Big)^2, \tag{5.3}$$

where $n_{g,t}$ denotes the size of group $g \in \{1, 2\}$ with label $t \in \{-1, 1\}$ in the sample. Intuitively, hybrid fairness requires both positive and both negatively labeled cross pairs to be treated similarly in average over the two groups. Some compensation might occur, but only amongst instances with the same label: over-valuing positive instances from group 1 can mitigate over-valuing positive instances from group 2, but does not mitigate under-valuing negative instances from group 1. These two terms can be weighted differently for applications where the treatment of positive (or negative) instances are more important. See Sections 5.3 and 5.4 for more details.

5.2.2. Discussion of Our Notions of Fairness

We now discuss several salient features of our fairness notions.

Why are these fairness notion different from accuracy? All of our fairness penalties are small for any perfect regressor (any $\mathbf{w}$ such that $\mathbf{w}^\top \mathbf{x} = y$ for all $(\mathbf{x}, y) \sim \mathcal{P}$): for a similar cross pair, $y_i \approx y_j$ and also $\mathbf{w}^\top \mathbf{x}_i \approx \mathbf{w}^\top \mathbf{x}_j$ for a perfect regressor $\mathbf{w}$. Our fairness regularizers might then be interpreted as an unusual proxy for standard accuracy rather than as fairness notions. However, perfect regressors almost never exist in practice; and between two models with similar accuracy, these definitions bias a learning procedure towards those which have similar treatment of similarly labeled instances from different groups.

What minimizes these penalties? We note that any constant regressor (any $\mathbf{w}$ such that $\mathbf{w}^\top \mathbf{x} = c$ for all $\mathbf{x} \in \mathcal{X}$ and some $c \in \mathbb{R}$) exactly minimizes all of our fairness regularizers. As we seen empirically, this implies that as the fairness regularization factor λ increases, we transition from an unfair model with minimum accuracy loss to a constant and perfectly fair model, whose accuracy is trivially the best any constant model can achieve.

5.3. Related Work

Recent work has shown that different fairness notions are often mutually exclusive [24, 37, 62]. Unsurprisingly then, different fairness notions have corresponded to different algorithms and optimization frameworks. Previously introduced fairness notions have generally split along several axis: classification vs. regression and individual vs. group fairness and disparate treatment. Most of previous work has focused on classification, despite the ubiquity of regression in real world applications with fairness concerns.

In classification, one line of work aims to achieve the group fairness notion known as *statistical parity*, i.e. to avoid disparate impact (see e.g. [1, 19, 35, 36, 44, 53–56, 73, 78]). Statistical parity requires a predictor to predict each label at similar rates across different groups. This definition can be at odds with accuracy especially when the two groups are inherently different. Hardt et al. [45] introduced a new notion of group fairness called *equality of odds*, partially to alleviate this friction, and partially arguing that equality of odds more accurately captured what it would mean for a classifier to be equally "good" for two groups. Equality of odds has a very intuitive interpretation for classification: it requires similarity of misclassification rates across groups (rather than forcing the marginal classification rates in the two groups to coincide). Optimizing for accuracy subject to an equality of odds constraint was recently shown to be NP-hard [88]; work following this result presented efficient heuristics for the problem [88, 90]. We also study fairness definitions which we can implement efficiently but our interest is in studying the trade-off between fairness and accuracy and the relationship between different notions of fairness.

Although equality of odds is also defined for regression, it is very difficult to determine empirically whether a regressor's output is conditionally independent of the protected attribute (conditioned on the true label), as each true label may be seen only once.

Calders et al. [20] introduced the study of statistical parity's analog in regression settings, (called *equal means* and *balanced residuals*). Johnson et al. [51] also studied fairness for regression problems and formalized several notions for *impartial estimates* based on the causal relationship between *sensitive attributes*, *legitimate attributes*, *suspect attributes* and the label. Both groups consider group fairness; our group fairness notion differs from these as we incorporate the similarity of pairs (through the function d) in our definition though the specific choice of $d(y, y') = c$ for some $c \in \mathbb{R}$ and all y and y' would recover equal means.

To achieve any of these fairness notions, one needs to decide whether or not to allow for *disparate treatment* (allowing for different treatment or models for different groups), and where in the learning process to enforce fairness: preprocessing of data (e.g. [53]); inprocessing, during the training of a model as either a constraint or incorporated into the objective function (e.g. [36, 56]); or postprocessing, where data is labeled by some black-box model and then relabeled as a function only of the original labels (e.g. [45]). Our approach in this paper falls into the in-processing category, by encoding fairness as a regularizer (an approach previously studied in e.g. [56, 90]). We differ from previous work in several aspects by primarily focusing on regression, and that our family of fairness measures draws inspiration from the idea that *similar* instances should be treated *similarly* [30, 91].

5.4. A Comparative Empirical Case Study

We now describe an empirical case study in which we apply our regularization framework to six different datasets in which fairness is a central concern. These datasets include cases in which the observed labels are real-valued, and cases in which they are binary-valued. For the real-valued datasets, we apply standard linear regression with our various fairness regularizers. For the binary-valued datasets, we apply logistic regression, again along with

fairness regularizers. For datasets with real-valued targets we normalized the inputs and outputs to be zero mean and unit variance, and we set the cross-group fairness weights as $d(y_i, y_j) = e^{-(y_i - y_j)^2}$; for datasets with binary targets we set $d(y_i, y_j) = \mathbb{1}[y_i = y_j]$.

For each dataset S, our framework requires that we solve optimization problems of the form $\min_{\mathbf{w}} \ell(\mathbf{w}, S) + \lambda f(\mathbf{w}, S) + \gamma ||\mathbf{w}||_2$ for variable values of λ, where $\ell(\mathbf{w}, S)$ is either MSE (linear regression) or the logistic regression loss. For each λ we picked γ as a function of this λ by cross validation. All optimization problems are solved using the CVX solver in Matlab (for real-valued datasets) or python (for binary-valued datasets). Furthermore, all the results are reported using 10-fold cross validation.

The datasets themselves are summarized in Table 1, where we specify the size and dimensionality of each, along with the "protected" feature (race or gender) that thus defines the subgroups across which we apply our fairness criteria. The datasets vary considerably in the number of observations, their dimensionality, and the relative size of the minority subgroup.

The *Adult* dataset [66, 69] from the UC Irvine Repository contains 1994 Census data, and the goal is to predict whether the income of an individual in the dataset is more than 50K per year or not. The sensitive or protected attribute is gender. The *Communities and Crime* dataset [69] includes features relevant to per capita violent crime rates in different communities in the United States, and the goal is to predict this crime rate; race is the protected variable. The *COMPAS* dataset contains data from Broward County, Florida, originally compiled by ProPublica [6], in which the goal is to predict whether a convicted individual would commit a violent crime in the following two years or not. The protected attribute is race, and the data was filtered in a fashion similar to that of Corbett-Davies et al. [25]. The *Default* dataset [69, 89] contains data from Taiwanese credit card users, and the goal is to predict whether an individual will default on payments. The protected attribute is gender. The *Law School* dataset consists of the records of law students who went on to take the bar exam. The goal is to predict whether a student will pass the exam based on features such as LSAT score and undergraduate GPA. The protected attribute is

gender. The *Sentencing* dataset contains information from a state department of corrections regarding inmates in 2010. The goal is to predict the sentence length given by the judge based on factors such as previous criminal records and the crimes for which the conviction was obtained. The protected attribute is gender.

Data Set	**Type**	n	d	**Minority** n	**Protected**
Adult	logistic regression	32561	14	10771	gender
Communities and Crime	linear	1994	128	227	race
COMPAS	logistic regression	3373	19	1455	race
Default	logistic regression	30000	24	11888	gender
Law School	logistic regression	27478	36	12079	gender
Sentencing	linear	5969	17	385	gender

Table 1: Summary of datasets. Type indicates whether regression is logistic or linear; n is total number of data points; d is dimensionality; Minority n is the number of data points in the smaller population; Protected indicates which feature is protected or fairness-sensitive.

5.4.1. Accuracy-Fairness Efficient Frontiers

We begin by examining the efficient frontier of accuracy vs. fairness for the six datasets. These curves are shown in Figure 11, and are obtained by varying the weight λ on the fairness regularizer, and for each value of λ finding the model which minimizes the associated regularized loss function. For the logistic regression cases, we extract probabilities from the learned model $\mathbf{w}$ as $\Pr[y_i = 1] = \exp(\mathbf{w}^\top x_i)/(1 + \exp(\mathbf{w}^\top x_i))$ and evaluate these probabilities as predictions for the binary labels using MSE. In all datasets, as λ increases, the models converge to the best constant predictor, which minimizes the fairness penalties.

Perhaps the most striking aspect of Figure 11 is the great diversity of tradeoffs across different datasets and different fairness regularizers. For instance, if we examine the individual fairness regularizer, on four of the datasets (Adult, Communities and Crime, Law School and Sentencing), the curvature is relatively mild and constant — there is an approximately fixed rate at which fairness can be traded for accuracy. In contrast, on COMPAS and Default, fairness loss can be reduced almost for "free" until some small threshold value, at which point the accuracy cost increases dramatically. Similar comments can be made

regarding hybrid fairness in the logistic regression cases.

Individual fairness appears to be strictly more costly than group fairness for the entire regime between the extremes of $\lambda = 0$ and $\lambda \to \infty$ for a majority of these datasets (with the exception of COMPAS and Default datasets). In COMPAS and Default, for small amounts of unfairness, group unfairness may be more costly than individual unfairness.

Perhaps surprisingly, building separate models for each population barely improves the tradeoff (and in some cases, hurts the tradeoff for some values of λ) across almost all datasets and fairness regularizers. This suggests that the academic discussion about whether to allow disparate treatment (as explicitly allowed in e.g. [30, 52]) – i.e. whether sensitive attributes such as race and gender should be "forbidden" vs. used in building more accurate models for each subpopulation, is perhaps less consequential than expected (at least on these datasets and using linear or logistic regression models).

Note that in theory group fairness is strictly less costly than individual fairness for any particular model $\mathbf{w}$ (by Jensen's inequality), and using separate models (one for each group) should strictly improve the fairness/accuracy trade-off for any of these notions of fairness. However, both group fairness and separate models are more prone to overfitting (than individual fairness and single model), and hence a larger ℓ_2-regularization parameter γ tends to be selected in cross validation in these settings. This is a surprising interaction between the strength of the fairness penalty and the generalization ability of the model, and results in group fairness sometimes having a more severe tradeoff with accuracy when compared to individual fairness, and separate models having little benefit out of sample, although they can appear to have a large effect in-sample (because of the effects of over-fitting).

5.4.2. Price of Fairness

The efficient fairness/accuracy frontiers pictured in Figure 11 can be compared across data sets in a qualitative sense — e.g. to see that in some datasets, the fairness penalty can be substantially decreased with little cost to accuracy. However, they are difficult to compare

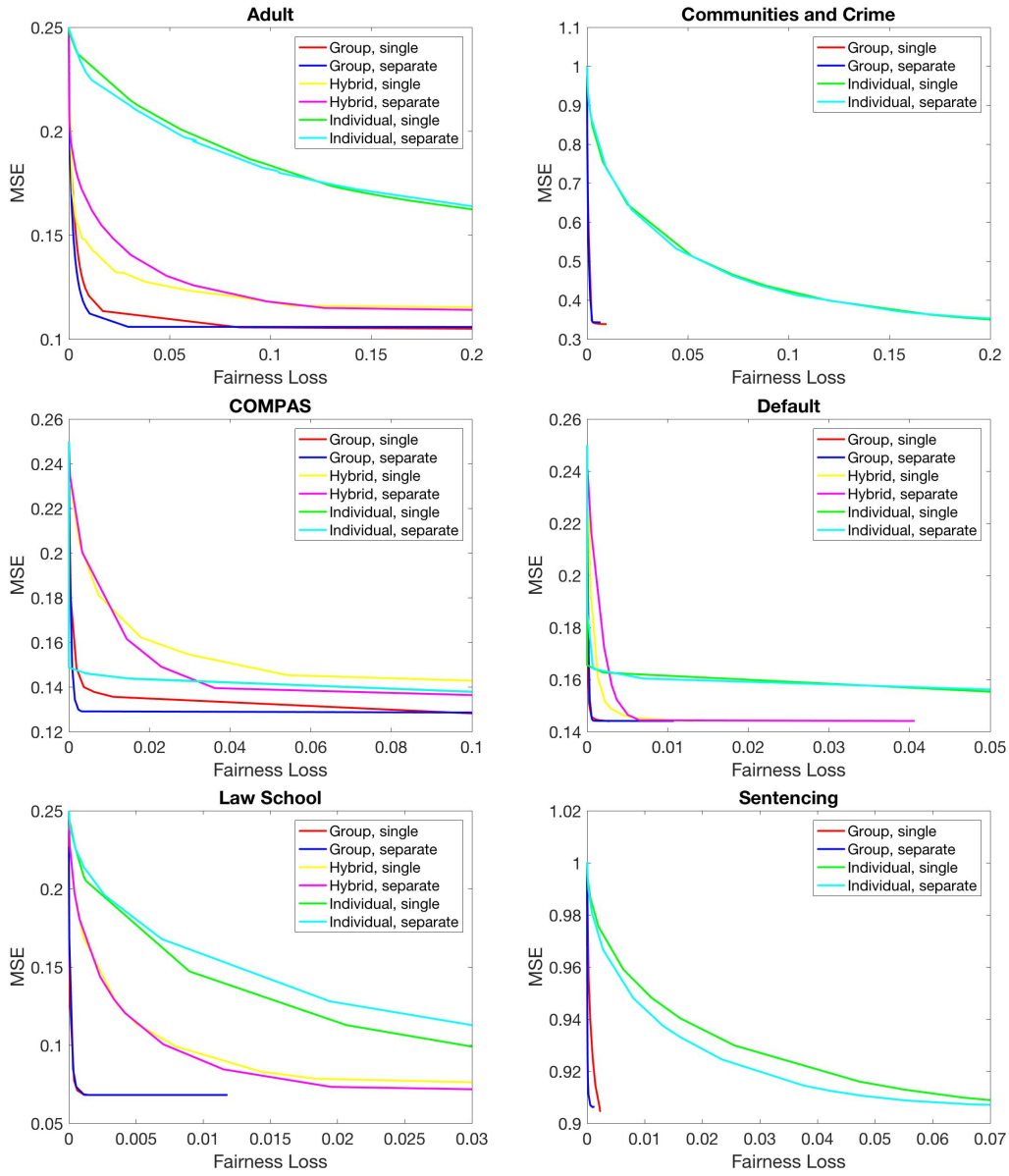


Figure 11: Efficient frontiers of accuracy vs. fairness for each dataset. For datasets with binary-valued targets (logistic regression), we consider three fairness notions (group, individual and hybrid), and for each examine building a single model or separate models for each group, yielding a total of six curves. For real-valued targets (linear regression), we consider two fairness notions (group and individual), and again single or separate models, yielding a total of four curves.

quantitatively as the scale of the fairness loss differs substantially from data set to data set. We next give a cross-dataset comparison using a measure we call *Price of Fairness* which has the effect of normalizing the fairness loss across data sets to lie on the same scale.

For a given data set and regression type (linear or logistic), let $\mathbf{w}^*$ be the optimal model absent any fairness penalty (i.e. the empirical risk minimizer when the fairness "regularization" weight $\lambda = 0$). This model will suffer some fairness penalty: it represents the "maximally unfair" point on the fairness/accuracy frontiers from Figure 11. For each dataset, we will fix a normalization such that this fairness penalty is rescaled to be 1, and ask for the cost (in terms of the relative increase in mean squared error) of constraining our predictor to have fairness penalty $\alpha \leq 1$. Equivalently, this is measuring the relative increase in MSE that results from constraining a predictor to have fairness penalty that is no more than *an α fraction of the fairness penalty of the unconstrained optimal predictor.*

More formally, let $\mathbf{w}^* = \arg\min_{\mathbf{w}} \ell_{\mathcal{P}}(\mathbf{w})$. For any value of $\alpha \in [0, 1]$ we define the *price of fairness* (PoF) as follows:

$$\text{PoF}(\alpha) = \frac{\min_{\mathbf{w}} \ell_{\mathcal{P}}(\mathbf{w}) \text{ subject to } f_{\mathcal{P}}(\mathbf{w}) \leq \alpha f_{\mathcal{P}}(\mathbf{w}^*)}{\ell_{\mathcal{P}}(\mathbf{w}^*)}.$$

Note that by definition, $PoF(\alpha) \geq 1$, $PoF(1) = 1$, and that $PoF(\alpha)$ increases monotonically as α decreases. Larger values represent more severe costs for imposing fairness constraints that ask that the measure of unfairness be small relative to the unconstrained optimum. It is important to note that because this measure asks for the cost of *relative* improvements over the unconstrained optimum, it can be, for example, that the PoF for one fairness penalty case is larger than for another, even if the *absolute* fairness loss for both the numerator and the denominator is smaller in the second case. With this observation in mind, we can move to the empirical findings.

Figure 12 displays the PoF on each of the 6 datasets we study, for each fairness regularizer (individual, hybrid, and group), and for the single and separate model case. Even when normalized on a common scale, we continue to see the diversity across datasets that was apparent in Figure 11. For some datasets (e.g. COMPAS and Sentencing), increasing the fairness constraint by decreasing α has only a mild cost in terms of error. For others (e.g.

Communities and Crime, and Law School), the cost increases steadily as we decrease α.

Next, we observe that with this normalization, although the difference between separate and single models remains small on most datasets, on two datasets, differences emerge. In the Law School dataset, restricting to a single model leads to a significantly higher PoF when considering the group fairness metric, compared to allowing separate models. In contrast, on the Adult dataset, restricting to a single model substantially *reduces* the PoF when considering the individual fairness metric.

Finally, this normalization allows us to observe variation across fairness penalties in the *rate of change* in the PoF as α is decreased. In some datasets (e.g. Communities and Crime, and Sentencing), the PoF changes in lock-step across all measures of unfairness. However, for others (e.g. Default), the PoF increases substantially with α when we consider group or hybrid fairness measures, but is much more stable for individual fairness.

5.5. Conclusions

The use of a complexity regularizer to control overfitting is both standard and well-understood in machine learning. While the use of such a regularizer introduces a trade-off — goodness of fit vs. model complexity — it does not introduce a *tension*, because regularization is always in service of improving generalization, and not a goal in its own right.

In contrast, in this work we have studied a variety of *fairness* regularizers for regression problems, and applied them to data sets in which fairness is not subservient to generalization, but is instead a first-order consideration. Our empirical study has demonstrated that the choice of fairness regularizer (group, individual, hybrid, or other) and the particular data set can have qualitative effects on the trade-off between accuracy and fairness. Combined with recent theoretical results [24, 37, 62] that also highlight the incompatibility of various fairness measures, our results highlight the care that must be taken by practitioners in defining the type of fairness they care about for their particular application, and in determining the appropriate balance between predictive accuracy and fairness.

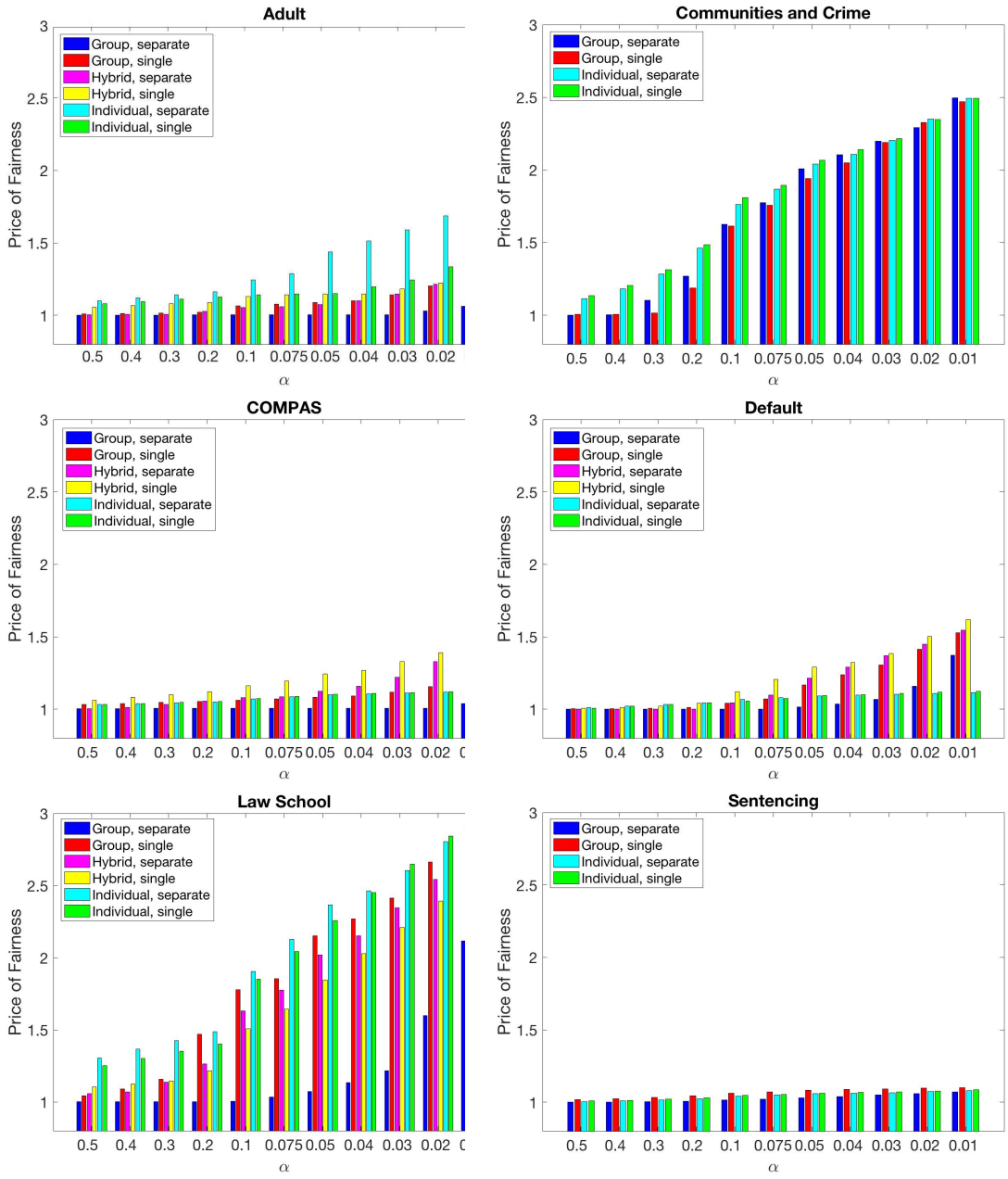


Figure 12: The PoF across data sets, for each type of fairness regularizer, in both the single and separate model case.

Chapter 6

Fairness in Allocations Problems

6.1. Introduction

The bulk of the literature on algorithmic fairness has focused on classification and regression problems (see e.g. [12, 13, 20, 23, 25, 29, 30, 45, 48, 52, 63, 71, 88, 90, 91] for a collection of recent work), but fairness concerns also arise naturally in many resource allocation settings. Informally, a resource allocation problem is one in which there is a limited supply of some *resource* to be distributed across multiple groups with differing needs. Resource allocation problems arise in financial applications (e.g. allocating loans), disaster response (allocating aid), and many other domains — but the primary example that we will focus on in this paper is policing. In the predictive policing problem, the resource to be distributed is police officers, which can be dispatched to different districts. Each district has a different crime distribution, and the goal (absent additional fairness constraints) might be to maximize the number of crimes caught. *We understand that policing has many goals besides simply apprehending criminals, including preventing crimes in the first place, fostering healthy community relations, and generally promoting public safety. But for concreteness and simplicity we consider the limited objective of apprehending criminals.*

Of course, fairness concerns abound in this setting, and recent work (see e.g. [32, 33, 72]) has highlighted the extent to which algorithmic allocation might exacerbate those concerns. For example, Lum and Isaac [72] show that if predictive policing algorithms such as PredPol are trained using past arrest data to predict future crime, then pernicious feedback loops can arise, which misestimate the true crime rates in certain districts, leading to an overallocation of police. Since the communities that Lum and Isaac [72] showed to be overpoliced on a relative basis were primarily poor and minority, this is especially concerning from a fairness perspective. In this work, we study algorithms that avoid this kind of under-exploration

and incorporate quantitative fairness constraints.

In the predictive policing setting, Ensign et al. [32] implicitly consider an allocation to be *fair* if police are allocated across districts in direct proportion to the district's crime rate; generally extended, this definition asks that units of a resource are allocated according to the group's share of the total candidates for that resource. In our work, we study a different notion of allocative fairness that has a similar motivation to the notion of *equality of opportunity* proposed by Hardt et al. [45] in classification settings. Informally speaking, it asks that the probability that a candidate for a resource be allocated a resource be independent of his group membership. In the predictive policing setting, it asks that conditional on committing a crime, the probability that someone is apprehended should not depend on the district in which they commit the crime.

To illustrate that our notions of fairness do not depend on whether individuals would prefer to receive or not receive the resource, we highlight another setting in which allocative fairness is a natural concern: hiring. Suppose a company wishes to recruit machine learning programmers by advertising on a social media platform. Many such platforms offer the ability to advertise to different demographics of users and charge by the number of times the advertisement is shown to different users (i.e., the number of *impressions*); a fixed advertising budget can then be viewed as a number of impressions to allocate. Depending on how well the platform can identify programmers within each demographic, the ad may be shown to a higher or lower number of programmers. In this setting, our notion of allocative fairness asks that the probability a programmer is exposed to the hiring ad (and thus, receives the opportunity to apply for a job) does not depend on the programmer's demographic, and the allocation problem is to maximize the number of programmers reached via the choice of impressions across each demographic, subject to fairness constraints.

6.1.1. Our Results

To define the extent to which an allocation satisfies our fairness constraint, we must model the specific mechanism by which resources deployed to a particular group reach their intended targets. We study two such *discovery models*, and we view the explicit framing of this modeling step as one of the contributions of our work; the implications of a fairness constraint depend strongly on the details of the discovery model, and specifying such a model is an important step in making one's assumptions transparent.

We study two discovery models, which capture two extremes of targeting ability. In the *random* discovery model, regardless of the number of units allocated to a given group, all individuals within that group are equally likely to be assigned a unit, regardless of whether they are a candidate for the resource or not. In other words, the probability that a candidate receives a resource is equal to the ratio of the number of units of the resource assigned to his group to the size of his group (*independent* of the number of candidates in the group).

At the other extreme, in the *precision* discovery model, units of the resource are given only to candidates within a group, as long as there is sufficient supply of the resource. That is, the probability that a candidate receives a resource is equal to the ratio of the number of units of the resource assigned to his group to the number of *candidates* in his group.

In the policing setting, these models can be viewed as two extremes of police targeting ability for an intervention like *stop-and-frisk*. In the random model, police are viewed as stopping people uniformly at random. In the precision model, police have the omniscient ability to identify individuals with contraband, and stop only them.

These discovery models have different implications for fairness. In the random model, fairness constrains resources to be distributed in amounts proportional to group sizes, regardless of the distribution of candidates, and so is uninteresting from a learning perspective. On the other hand, the precision model yields an interesting fairness-constrained learning prob-

lem when the distribution of the number of candidates in each group must be learned via observation, and what counts as a 'fair' allocation depends greatly on these distributions.

We study learning in a censored feedback setting: each round, the algorithm can choose a feasible deployment of resources across groups. Then the number of candidates for the current round in each group is drawn from a fixed, but unknown group-dependent distribution (which might be not be independent from the distributions in other groups). The algorithm does not observe the number of candidates present in each group, but only the number of candidates that received the resource. In the policing setting, this corresponds to the algorithm being able to observe the number of arrests, but not the actual number of crimes. Thus, the extent to which the algorithm can learn about the distribution in a particular group is limited by the number of resources it deploys there. The goal of the algorithm is to converge to an optimal fairness-constrained allocation, where here both the objective value and the constraints imposed on it depend on the unknown distributions.

One trivial solution to the learning problem is to sequentially deploy *all* resources to each group in turn for a sufficient amount of time to accurately learn the candidate distributions. This would reduce the learning problem to an offline constrained optimization problem, which we show can be efficiently solved by a greedy algorithm. But this algorithm is unreasonable: it has a large exploration phase in which it uses nonsensical deployments, vastly overallocating to some groups and underallocating to others. A more natural approach is a greedy-style algorithm which at each round uses its current best-guess estimate for the distribution in each group and deploys an optimal fairness-constrained allocation according to these estimates. Unfortunately, as we show, if one makes no assumptions on the underlying distributions, any algorithm that has a guarantee of converging to a fair allocation must behave like the trivial one, deploying vast numbers of resources to each group in turn.

This impossibility result motivates us to consider the learning problem in which the unknown distributions are from a known parametric family. The natural greedy algorithm uses an optimal fair deployment at each round given the maximum likelihood estimates of candidate

distributions given its (censored) observations so far; for concreteness, we consider the case of the Poisson distribution, and show that it converges to an optimal fair allocation, but our analysis generalizes for any single-parameter Lipschitz-continuous family of distributions.

Finally, we conduct an empirical evaluation of our algorithm on the *Philadelphia Crime Incidents* dataset, which records all crimes reported to the Philadelphia Police Department's INCT system between 2006 and 2016. We verify that the crime distributions in each district are in fact well-approximated by Poisson distributions, and that our algorithm converges quickly to an optimal fair allocation (as measured according to the empirical crime distributions in the dataset). We also systematically evaluate the *Price of Fairness*, and plot the Pareto curves that trade off the number of crimes caught versus the slack allowed in our fairness constraint, for different sizes of police force, on this dataset. For the random discovery model, we prove worst-case bounds on the Price of Fairness.

6.1.2. Further Related Work

Our precision discovery model is inspired by and has technical connections to Ganchev et al. [39], which models the *dark pool problem* from quantitative finance, in which a trader wishes to execute a specified number of trades across a set of exchanges of unknown but independently distributed liquidity. In Ganchev et al. [39], the authors design an optimal allocation algorithm under the censored feedback of the precision model. It is straightforward to map their setting onto ours, but they assume independence between different exchanges, while the candidate distributions in our setting need not be independent. Regardless, we show that their allocation algorithm can be used to compute an optimal allocation (ignoring fairness) even when the independence assumption is relaxed (see Remark 1). Later, Agarwal et al. [2] extend the dark pool problem to an adversarial (rather than distributional) setting. This is quite closely related to the work of Ensign et al. [33] who also consider the precision model (under a different name) in an adversarial predictive policing setting. They provide no-regret algorithms for this setting by reducing the problem to learning in a partial monitoring environment. Since their setting is equivalent to that of Agarwal et al.

[2], the algorithms in Agarwal et al. [2] can be directly applied to their problem.

Our desire to study the natural greedy algorithm rather than an algorithm which uses "unreasonable" allocations during an exploration phase is an instance of a general concern about exploration in fairness-related problems [14]. Recent works have studied the performance of greedy algorithms in different settings for this reason [11, 57, 80].

Lastly, the term *fair allocation* appears in the *fair division* literature (see e.g. [79]), but that body of work is technically quite distinct from the problem we study here.

6.2. Setting

We study an *allocator* who has $\mathcal{V}$ units of a resource and is tasked with distributing them across a population partitioned into $\mathcal{G}$ groups. Each group is divided into *candidates*, who are the individuals the allocator would like to receive the resource, and *non-candidates*, who are the remaining individuals. We let m_i denote the total number of individuals in group i. The number of candidates c_i in group i is a random variable drawn from a fixed but unknown distribution $\mathcal{C}_i$ called the *(marginal) candidate distribution*. We do not make any assumptions about the relationship between the candidate distributions across different groups and in particular these distributions need not be independent. We use M to denote the total size of all groups. An allocation $\mathbf{v} = (v_1, \ldots, v_{\mathcal{G}})$ is a partitioning of these $\mathcal{V}$ units, where $v_i \in \{0, \ldots, \mathcal{V}\}$ denotes the units of resources allocated to group i. Every allocation is bound by a *feasibility* constraint which requires that $\Sigma_{i \in [\mathcal{G}]} v_i \leq \mathcal{V}$.

A *discovery model* $\text{disc}(v_i, c_i)$ is a (possibly randomized) function mapping the number of units v_i allocated to group i and the number of candidates c_i in group i to the number of candidates discovered in group i. In the learning setting, upon fixing an allocation $\mathbf{v}$, the learner will get to observe (a realization of) $\text{disc}(v_i, c_i)$ for the realized value of c_i for each group i. Fixing an allocation $\mathbf{v}$, a discovery model $\text{disc}(\cdot)$ and candidate distributions for all groups $\mathcal{C} = \{\mathcal{C}_i : i \in [\mathcal{G}]\}$, we define the total expected number of discovered candidates,

$\chi(\mathbf{v}, \text{disc}(\cdot), \mathcal{C})$, as

$$\chi\left(\mathbf{v}, \text{disc}(\cdot), \mathcal{C}\right) = \sum_{i \in [\mathcal{G}]} \mathbb{E}_{c_i \sim \mathcal{C}_i}\big[\,\text{disc}(v_i, c_i)\big], \tag{6.1}$$

where the expectation is taken over $\mathcal{C}_i$ and any randomization in the discovery model $\text{disc}(\cdot)$. When the discovery model and the candidate distributions are fixed, we will simply write $\chi(\mathbf{v})$ for brevity. We also use the total expected number of discovered candidates and *(expected) utility* exchangeably. We refer to an allocation that maximizes the expected utility over all feasible allocations as an *optimal allocation* and denote it by $\mathbf{w}^*$.

6.2.1. Allocative Fairness

For the purposes of this paper, we say that an allocation is *fair* if it satisfies *approximate equality of candidate discovery probability* across groups. We call this *discovery probability* for brevity. This formalizes the intuition that it is unfair if candidates in one group have an inherently higher probability of receiving the resource than candidates in another. Formally, we define our notion of *allocative fairness* as follows.

Definition 15. *Fix a discovery model* $\text{disc}(\cdot)$ *and the candidate distributions* $\mathcal{C}$*. For an allocation* $\boldsymbol{v}$*, let*

$$f_i\left(v_i, \text{disc}(\cdot), \mathcal{C}_i\right) = \mathbb{E}_{c_i \sim \mathcal{C}_i}\left[\frac{\text{disc}\left(v_i, c_i\right)}{c_i}\right],$$

denote the expected probability that a random candidate from group i *receives a unit of the resource at allocation* $\boldsymbol{v}$ *(i.e. the discovery probability in group* i*). Then for any* $\alpha \in [0, 1]$*,* $\boldsymbol{v}$ *is* α*-fair if for all pairs of groups* i *and* j *we have that*

$$\left| f_i\left(v_i, \text{disc}(\cdot), \mathcal{C}_i\right) - f_j\left(v_j, \text{disc}(\cdot), \mathcal{C}_j\right) \right| \leq \alpha.$$

When it is clear from the context, for brevity, we write $f_i(v_i)$ for the discovery probability in group i. We emphasize that this definition (1) depends crucially on the chosen discovery model, and (2) requires nothing about the treatment of non-candidates. We think of this as a *minimal* definition of fairness, in that one might want to further constrain the treatment

of non-candidates — but we do not consider that extension.

Since discovery probabilities $f_i(v_i)$ and $f_j(v_j)$ are in $[0, 1]$, the absolute value of their difference is in $[0, 1]$. By setting $\alpha = 1$ we impose no fairness constraints whatsoever on the allocations, and by setting $\alpha = 0$ we require *exact* fairness.

We refer to an allocation $\mathbf{v}$ that maximizes $\chi(\mathbf{v})$ subject to α-fairness and the feasibility constraint as an *optimal α-fair allocation* and denote it by $\mathbf{w}^\alpha$. $\chi(\mathbf{w}^\alpha)$ is a non-increasing in α, since as α diminishes, the utility maximization problem becomes more constrained.

Remark 1. *We note that both the utility and discovery probabilities can be written solely in terms of the* marginal *candidate distributions in each of the groups, even when these distributions are not independent. This is because we have (implicitly) assumed that the number of candidates discovered in a group depends only on the number of candidates in the group and the allocation to that group, regardless of the allocations to and the number of candidates in other groups. This assumption together with the linearity of expectation allows us to write the expected utility as in the right hand side of Equation 6.1.*

6.3. The Precision Discovery Model

We begin by describing the *precision model* of discovery. Allocating v_i units to group i in the precision model results in the discovery of $\text{disc}(c_i, v_i) \triangleq \min(c_i, v_i)$ candidates. This models the ability to perfectly discover and reach candidates in a group with resources deployed to that group, limited only by the number of deployed resources and the number of candidates present.

The precision model results in *censored* observations that have a particularly intuitive form. Recall that in general, a learning algorithm at each round gets to choose an allocation $\mathbf{v}$ and then observes $\text{disc}(v_i, c_i)$ for each group i. In the precision model, this results in the following kind of observation: when v_i is larger than c_i, the allocator learns the number of candidates c_i present on that day exactly. We refer to this kind of feedback as an *uncensored observation*. When v_i is smaller than c_i, all the allocator learns is that the

number of candidates is *at least* v_i. We call this a *censored observation*.

The rest of this section is organized as follows. In Sections 6.3.1 and 6.3.2 we characterize optimal and optimal fair allocations for the precision model when the candidate distributions are known. In Section 6.3.3 we focus on learning an optimal fair allocation when these distributions are unknown. We show that any learning algorithm that is guaranteed to find a fair allocation in the *worst case* over candidate distributions must have the undesirable property that at some point, it must allocate a vast number of its resources to each group individually. To bypass this hurdle, in Section 6.3.4 we show that when the candidate distributions have a parametric form, a natural greedy algorithm which always uses an optimal fair allocation for the current maximum likelihood estimates of the candidate distributions converges to an optimal fair allocation.

6.3.1. Optimal Allocation

We first describe how an optimal allocation (absent fairness constraints) can be computed efficiently when the candidate distributions $\mathcal{C}_i$ are known. In Ganchev et al. [39], the authors provide an algorithm for computing an optimal allocation when the distributions over the number of shares present in each dark pool are known and the trader wishes to maximize the expected number of traded shares. They assume that the distributions of shares across different dark pools are independent, but our formulation does not require this assumption of independence. Still, we can use the same algorithm as in Ganchev et al. [39] to compute an optimal allocation in our setting; this is because, as stated in Remark 1, the utility in both settings can be written solely in terms of the (marginal) candidate distributions even when the candidate distributions are not independent across groups. Here, we present the high level ideas of their algorithm in the language of our model.

Let $\mathcal{T}_i(c) = \Pr_{c_i \sim \mathcal{C}_i}[c_i \geq c]$ denote the probability that there are at least c candidates in group i. We refer to $\mathcal{T}_i(c)$ as the *tail probability of* $\mathcal{C}_i$ *at* c. Recall that the value of the

cumulative distribution function (CDF) of $\mathcal{C}_i$ at c is defined to be

$$\mathcal{F}_i(c) = \sum_{c' \leq c} \Pr_{c_i \sim \mathcal{C}_i} \left[c_i = c' \right].$$

So $\mathcal{T}_i(c)$ can be written in terms of CDF values as $\mathcal{T}_i(c) = 1 - \mathcal{F}_i(c-1)$.

The expected total number of candidates discovered by an allocation in the precision model can be written in terms of the tail probabilities of the candidate distributions i.e.

$$\chi(\mathbf{v}, \text{disc}(\cdot), \mathcal{C}) = \sum_{i \in [\mathcal{G}]} \mathbb{E}_{c_i \sim \mathcal{C}_i} \left[\min \left(v_i, c_i \right) \right] = \sum_{i \in [\mathcal{G}]} \sum_{c=1}^{v_i} \mathcal{T}_i(c).$$

Since the objective function is concave (as $\mathcal{T}_i(c)$ is a non-increasing function in c for all i), a greedy algorithm which iteratively allocates the next unit of the resource to a group in

$$\text{argmax}_{i \in [\mathcal{G}]} \left(\mathcal{T}_i \left(v_i^t + 1 \right) - \mathcal{T}_i \left(v_i^t \right) \right),$$

where v_i^t is the current allocation to group i in the t^{th} round achieves an optimal allocation.

6.3.2. Optimal Fair Allocation

We next show how to compute an optimal α-fair allocation in the precision model when the candidate distributions are known and do not need to be learned.

To build intuition for how the algorithm works, imagine that the group i has the highest discovery probability in $\mathbf{w}^\alpha$, and the allocation w_i^α to that group is somehow known to the algorithm ahead of time. The constraint of α-fairness then implies that the discovery probability for each other group j in $\mathbf{w}^\alpha$ must satisfy $f_j \in [f_i - \alpha, f_i]$. This in turn implies upper and lower bounds on the feasible allocations w_j^α to group j. The algorithm is then simply a constrained greedy algorithm: subject to these implied constraints, it iteratively allocates units so as to maximize their marginal probability of reaching another candidate. Since the group i maximizing the discovery probability in $\mathbf{w}^\alpha$ and the corresponding allocation w_i^α

are not known ahead of time, the algorithm simply iterates through each possible choice of i. Pseudocode is given in Algorithm 1.

Algorithm 1 Computing an optimal fair allocation in the precision model

Input: α, $\mathcal{C}$ and $\mathcal{V}$.
Output: An optimal α-fair allocation $\mathbf{w}^\alpha$.

```
w^α ← 0⃗.                                   ▷ Initialize the output.
χ_max ← 0.                                  ▷ Keep track of the utility of the output.
for i ∈ [G] do                              ▷ Guess for group with the highest probability of discovery.
    v ← 0⃗.
    for v_i ∈ {0, ... V} do                 ▷ Guess for the allocation to that group.
        Set v_i in v and compute f_i(v_i).
        ub_i ← v_i.                         ▷ Upper bound on allocation to group i.
        lb_i ← v_i.                         ▷ Lower bound on allocation to group i.
        for j ≠ i, j ∈ [G] do               ▷ Upper and lower bounds for other groups.
            Update lb_j and ub_j using f_i(v_i), α and C_j.
            v_j ← lb_j.                     ▷ Assign the lower bound allocation to group j.
        if Σ_{i∈[G]} v_i > V then
            continue.                       ▷ Allocation is not feasible.
        V_r = V − Σ_{i∈[G]} v_i
        for t = 1, ..., V_r do              ▷ Allocate the remaining resources greedily while obeying fairness.
            j* ∈ argmax_{j∈[G]} (T_j(v_j + 1) − T_j(v_j)) s.t. v_j < ub_j.
            v_{j*} ← v_{j*} + 1.
        χ(v) = Σ_{i∈[G]} Σ_{c=1}^{v_i} T_i(c).   ▷ Compute the utility of v.
        if χ(v) > χ_max then                ▷ Update the best α-fair allocation found so far.
            χ_max ← χ(v).
            w^α ← v.
return w^α.
```

We prove that Algorithm 1 returns an optimal α-fair allocation in Theorem 10.

Theorem 10. *Algorithm 1 computes an optimal α-fair allocation for the precision model in time $O(\mathcal{G}\mathcal{V}(\mathcal{G}\mathcal{V} + M))$.*

6.3.3. Learning Fair Allocations Generally Requires Brute-Force Exploration

In Sections 6.3.1 and 6.3.2 we assumed the candidate distributions were known. When the candidate distributions are unknown, learning algorithms intending to converge to optimal α-fair allocations must learn a sufficient amount about the distributions in question to

certify the fairness of the allocation they finally output. Because learners must deal with feedback in the censored observation model, this places constraints on how they can proceed. As we show in Theorem 11, if candidate distributions are allowed to be worst-case, this will force a learner to engage in "brute-force exploration" — the iterative deployment of a large fraction of the resources to each subgroup in turn.

Theorem 11. *Define* $m^* = \max_{i \in [\mathcal{G}]} m_i$ *to be the size of the largest group and assume* $m_i > 6$ *for all* i *and* $\mathcal{G} \geq 2$. *Let* $\alpha \in [0, 1/(2m^*))$, $\delta \in (0, 1/2)$, *and* $\mathcal{A}$ *be any learning algorithm for the precision model which runs for a finite number of rounds and outputs an allocation. Suppose that there is some group* i *for which* $\mathcal{A}$ *has not allocated at least* $m_i/2$ *units for at least* $k \ln(1/\delta)/(\alpha m_i)$ *rounds upon termination, where* k *is an absolute constant. Then there exists a candidate distribution such that, with probability at least* δ, $\mathcal{A}$ *outputs an allocation that is not* α*-fair.*

Recall that we used m^* to denote the size of the largest group. When $m^* > 2\mathcal{V}$, then Theorem 11 implies that no algorithm can guarantee α-fairness for sufficiently small α. Moreover, even when $m^* \leq 2\mathcal{V}$, Theorem 11 shows that in general, if we want algorithms that have provable guarantees for *arbitrary* candidate distributions, it is impossible to avoid something akin to brute-force search (recall that there is a trivial algorithm which simply allocates *all* resources to each group in turn, for sufficiently many rounds to approximately learn the CDF of the candidate distribution, and then solves the offline problem). In the next section, we circumvent this by giving an algorithm with provable guarantees, assuming that the candidate distributions have a known parametric form.

6.3.4. Poisson Distributions and Convergence of the MLE

In this section, we assume that all the candidate distributions have a particular and known *parametric form* but that the parameters of the these distributions are not known to the allocator. Concretely, we assume that the candidate distribution for each group is Poisson (denoted by $\mathcal{C}(\lambda)$) and write $\boldsymbol{\lambda}^* = (\lambda_1^*, \ldots, \lambda_{\mathcal{G}}^*)$ for the true underlying parameters of the candidate distributions; this choice appears justified, at least in the predictive policing

application, as the candidate distributions in the Philadelphia Crime Incidents dataset are well-approximated by Poisson distributions (see Section 6.4 for further discussion). This assumption allows an algorithm to learn the tails of these distributions without needing to rely on brute-force search, thus circumventing the limitation given in Theorem 11. Indeed, we show that (a small variant of) the natural greedy algorithm incorporating these distributional assumptions converges to an optimal fair allocation.

For simplicity, we assume a parametric form on the marginal candidate distribution in each of the groups. We could have equivalently assumed that the candidates across groups are drawn from a multivariate Poisson distribution to highlight the (potential) correlation between candidates distributions. However, since for a given multivariate Poisson distribution the marginal distribution on each group is itself a Poisson distribution [46], we made our parametric assumption directly on these marginal distributions.

At a high level, in each round, our algorithm uses Algorithm 1 to calculate an optimal fair allocation with respect to the current maximum likelihood estimates of the group distributions; then, it uses the new observations it obtains from this allocation to refine these estimates for the next round. This is summarized in Algorithm 2. The algorithm differs from this pure greedy strategy in one respect, to overcome the following subtlety: there is a possibility that Algorithm 1, when operating on a preliminary estimate for the candidate distributions, will suggest sending zero units to some group, even when the optimal allocation for the true distributions sends some units to every group. Such a deployment would result in the algorithm receiving no feedback for the zero-allocated group that round. If this suggestion is followed and a lack of feedback is allowed to persist indefinitely, the algorithm's parameter estimate for the zero-allocated group will also stop updating — potentially at an incorrect value. In order to avoid this problem and continue making progress in learning, our algorithm chooses another allocation in this case. As we show, any allocation that allocates positive resources to all groups will suffice; in particular, our algorithm simply repeats the allocation from the previous round.

Algorithm 2 Learning an optimal fair allocation

Input: α, $\mathcal{V}$ and T (total number of rounds).
Output: An allocation $\mathbf{v}^{T+1}$ and estimates to parameters $\{\lambda_i^T\}$.

```
v^1 ← (⌊(V/G)⌋, ..., ⌊(V/G)⌋).                    ▷ Allocate uniformly.
for rounds t = 1, ..., T do
    if ∃i such that v_i^t == 0 then    ▷ Check whether every group is allocated a resource.
        v^t ← v^{t-1}.
    Observe o_i^t = min{v_i^t, c_i^t} for each group.
    for i = 1, ..., G do
        Update history h_i^{t+1} with o_i^t and v_i^t.
        λ̂_i^t ← argmax_{λ∈[λ_min,λ_max]} L̂(h_i^{t+1}, λ).          ▷ Solve the MLE.
    v^{t+1} ← Algorithm 1(α, {C(λ̂_i^t)}, V).    ▷ Compute an allocation for the next round.
return v^{T+1} and {λ_i^T}.
```

Algorithm 2 chooses an allocation at every round which is fair with respect to its estimates of the parameters of the candidate distributions; hence, asymptotic convergence of its output to an *optimal α-fair* allocation follows directly from the convergence of the estimates to true parameters. However, we seek a *finite sample* guarantee, as stated in Theorem 12.

Theorem 12. *Let $\epsilon, \delta > 0$. Suppose that the candidate distributions are Poisson distributions with unknown parameters in the vector $\boldsymbol{\lambda}^*$, where $\boldsymbol{\lambda}^*$ lies in the known interval $[\lambda_{\min}, \lambda_{\max}]^{\mathcal{G}}$. Suppose we run Algorithm 2 for $t > \tilde{\mathcal{O}}(\ln(\mathcal{G}/\delta)/(\eta(\epsilon))^2) \triangleq T_{\max}$ rounds, where $\eta(\cdot)$ is some distribution specific function to get an allocation $\hat{\boldsymbol{v}}$ and estimated parameters $\hat{\lambda}_i$ for all groups i. Then with probability at least $1-\delta$*

1. *For all i in $[\mathcal{G}]$, $|\hat{\lambda}_i - \lambda_i^*| \leq \epsilon$.*
2. *Let $D = \max_{i\in[\mathcal{G}]} D_{TV}(\mathcal{C}(\lambda_i^*), \mathcal{C}(\hat{\lambda}_i))$ where D_{TV} denotes the total variation distance between two distributions. Then $\hat{\boldsymbol{v}}$*
   - *is $(\alpha + 4D)$-fair.*
   - *has utility at most $4D\mathcal{G}\mathcal{V}$ smaller than the utility of an optimal $(\alpha - 4D)$-fair allocation i.e. $\chi(\hat{\boldsymbol{v}}) \geq \chi(\boldsymbol{w}^{\alpha-4D}) - 4D\mathcal{G}\mathcal{V}$.*

Remark 2. *Theorem 12 implies that in the limit, the allocation from Algorithm 2 will converge to an optimal α-fair allocation. As $t \to \infty$, $\hat{\lambda}_i \xrightarrow{p} \lambda_i^*$ for all i, meaning $D \to 0$ and*

more importantly, $\hat{\boldsymbol{v}}$ will be α-fair and optimal.

To prove Theorem 12, we first show that *any* sequence of allocations selected by Algorithm 2 will eventually recover the true parameters. There are two difficulties here: the first is that standard convergence results typically leverage the assumption of *independence*, which does not hold in this case as Algorithm 2 computes *adaptive* allocations which depend on the allocations in previous rounds; the second is the censoring of the observations. Despite these difficulties, we give quantifiable rates with which the estimates converge to the true parameters. Next, we show that computing an optimal α-fair allocation using the estimated parameters will result in an allocation that is $(\alpha+4D)$-fair with respect to the true candidate distributions where D denotes the maximum total variation distance between the true and estimated Poisson distributions across all groups. Finally, we show that this allocation also achieves a utility that is comparable to the utility of an optimal $(\alpha-4D)$-fair allocation.

Remark 3. *Although we assumed Poisson distributions in this section, all our results hold for any single-parameter Lipschitz-continuous distribution whose parameter is drawn from a compact set. However, the convergence rate of Theorem 12 depends on the quantity $\eta(\epsilon)$ which* depends *on the family of distributions used to model the candidate distributions.*

6.4. Experiments

We next apply our allocation and learning algorithms for the precision model to the Philadelphia Crime Incidents dataset, and complement the theoretical convergence guarantee of Algorithm 2 to an optimal fair allocation with empirical evidence suggesting fast convergence in practice. We also study the trade-off between fairness and utility in the dataset.

6.4.1. Experimental Design

The Philadelphia Crime Incidents dataset (https://www.opendataphilly.org/dataset/crime-incidents accessed 2018-05-16) contains all the crimes reported to the Police Department's INCT system between 2006 and 2016. The crimes are divided into two types. Type I crimes include violent offenses such as aggravated assault, rape, and arson among

others. Type II crimes include simple assault, prostitution, gambling and fraud. For simplicity, we aggregate all crime of both types, but in practice, a police department would of course treat different categories of crime differently. We note as a caveat that these crimes are *reported* and may not represent the entirety of *committed* crimes.

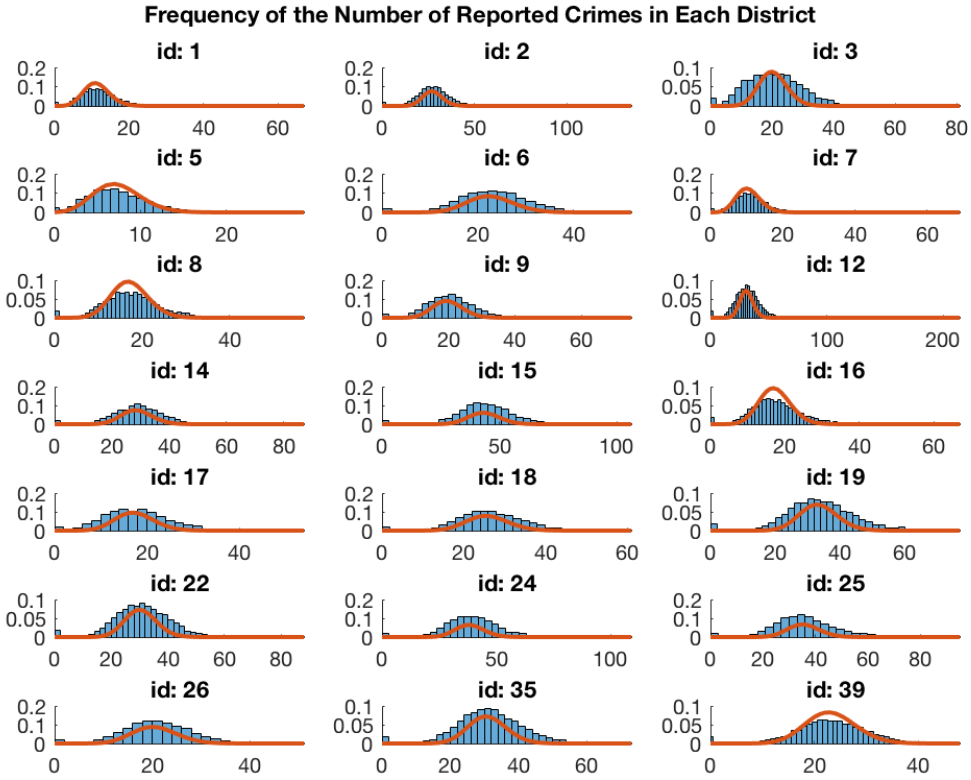


Figure 13: Frequencies of the number of reported crimes in each district in the Philadelphia Crime Incidents dataset. The red curves display the best Poisson fit to the data.

To create daily crime frequencies in Figure 13, we first calculate the daily counts of criminal incidents in each of the 21 geographical police districts in Philadelphia by grouping together all the crime reports with the same date; we then normalize these counts to get frequencies. Each subfigure in Figure 13 represents a police district. The horizontal axis of the subfigure corresponds to the number of reported incidents in a day and the vertical axis represents the frequency of each number on the horizontal axis. These frequencies approximate the true (marginal) distributions of the number of reported crimes in each of the districts in Philadelphia. Therefore, throughout this section we take these frequencies as the *ground truth* candidate distributions for the number of reported incidents in each of the districts.

Figure 13 shows that crime distributions in different districts can be quite different; e.g., the average number of daily reported incidents in District 15 is 43.5, which is much higher than the average of 11.35 in District 1. Despite these differences, each of the crime distributions can be approximated well by a Poisson distribution. The red curves overlayed in each subfigure correspond to the Poisson distribution obtained via maximum likelihood estimation on data from that district. Throughout, we refer to such distributions as the *best Poisson fit* to the data. districts as the resource to be distributed, the ground truth crime frequencies as candidate distributions, and aim to maximize the sum of the number of crimes discovered under the precision model of discovery.

6.4.2. Results

We can quantify the extent to which fairness degrades utility in the dataset through a notion we call *Price of Fairness* (PoF henceforth). In particular, given the ground truth crime distributions and the precision model of discovery, for a fairness level α, we define $\text{PoF}(\alpha) = \chi(\mathbf{w}^*)/\chi(\mathbf{w}^\alpha)$. The PoF is simply the ratio of the expected number of crimes discovered by an optimal allocation to the expected number of crimes discovered by an optimal α-fair allocation. Since $\chi(\mathbf{w}^*) \geq \chi(\mathbf{w}^\alpha)$ for all α, the PoF is at least one. Furthermore, the PoF is monotonically non-increasing in α. We can apply the algorithms given in Sections 6.3.1 and 6.3.2 respectively for computing optimal unconstrained, and optimal fair allocations with the with ground truth distributions as input and numerically compute the PoF. This is illustrated in Figure 14. The x axis corresponds to different α values and the y axis displays $1/\text{PoF}(\alpha)$. Each curve corresponds to a different number of total police officers denoted by $\mathcal{V}$. Because feasible allocations must be integral, there can sometimes be no feasible α-fair allocation for small α. Since the PoF in these cases is infinite we instead opt to display the inverse, 1/PoF, which is always bounded in $[0, 1]$. Higher values of inverse PoF are more desirable (see [27] for a theoretical study of PoF in the precision model).

Figure 14 shows a diverse set of utility/fairness trade-offs depending on the number of available police officers. It also illustrates that the cost of fairness is rather low in most regimes.

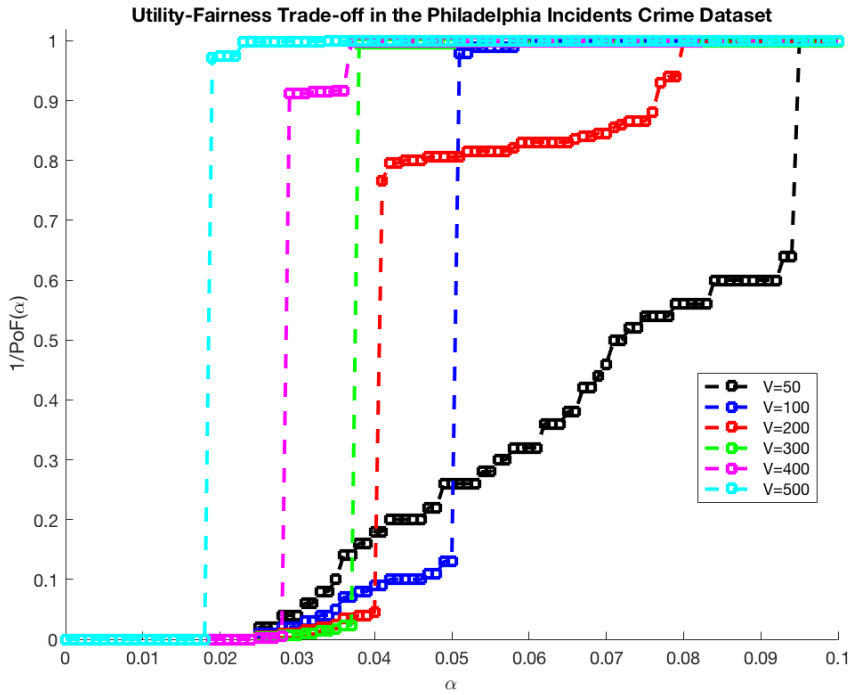


Figure 14: Inverse PoF plots for the Philadelphia Crime Incidents dataset. Smaller values indicate greater sacrifice in utility to meet the fairness constraint.

For example, in the worst case, with only 50 police officers (the black curve) (which is much smaller than the average number of daily reported crimes: 563.88), the inverse PoF is 1 for $\alpha \geq 0.1$, which corresponds to a 10% difference in the discovery probability across districts. When we increase the number of available police officers to 400 (the magenta curve), tolerating only a 4% difference in the discovery probability across districts is sufficient to guarantee no loss in the utility. Figure 14 also shows that for any fixed α, the inverse PoF(α) tends to increase as the number of police increases (i.e. the cost of fairness decreases). This captures the intuition that fairness becomes a less costly constraint when resources are in greater supply. Finally, we observe a thresholding phenomenon in Figure 14; in each curve, increasing α beyond a threshold will significantly increase the inverse PoF. This is due to discretization effects, since only integral allocations are feasible.

We next turn into analyzing the performance of Algorithm 2 in practice. We run the algorithm instantiated to fit Poisson distributions, but use observations from the ground

truth distribution at each round. As we have shown in Figure 13,the ground truth is well approximated by a Poisson distribution.

We measure the performance of Algorithm 2 as follows. First, we fix a police budget $\mathcal{V}$ and unfairness budget α and run Algorithm 2 for 2000 rounds using the dataset as the ground truth. That is, we simulate each round's crime count realizations in each of the districts as being sampled from the ground truth distributions, and return censored observations under the precision model to Algorithm 2 according to the algorithm's allocations and the drawn realizations. The algorithm returns an allocation after termination and we can measure the expected number of crimes discovered and fairness violation (the maximum difference in discovery probabilities over all pairs of districts) of the returned allocation using the ground truth distributions. Varying α while fixing $\mathcal{V}$ allows us to trace out the Pareto frontier of the utility/fairness trade-off for a fixed police budget. Similarly, for any fixed $\mathcal{V}$ and α, we can run Algorithm 1 (the offline algorithm for computing an optimal fair allocation) with the ground truth distributions as input and trace out a Pareto curve by varying α. We refer to these two Pareto curves by the *learned* and *optimal* Pareto curves, respectively. To measure the performance of Algorithm 2, we can compare these curves.

In Figure 15, each curve corresponds to a police budget. The x and y axes represent the expected number of crimes discovered and fairness violation for allocations on the Pareto frontier, respectively. In our simulations we varied α between 0 and 0.15. For each police budget $\mathcal{V}$, the 'x' s connected by the dashed lines show the learning Pareto frontier. Similarly, the circles connected by solid lines show the optimal Pareto frontier. We point out that while it is possible for the fairness violations in the learned Pareto curves to be higher than the level of α set as an input to Algorithm 2, the fairness violations in the optimal Pareto curves are always bounded by α.

The disparity between the optimal and learned Pareto curves are due to the fact that the learning algorithm has not yet fully converged. This can be attributed to the large number of censored observations received by Algorithm 2, which are significantly less informative than

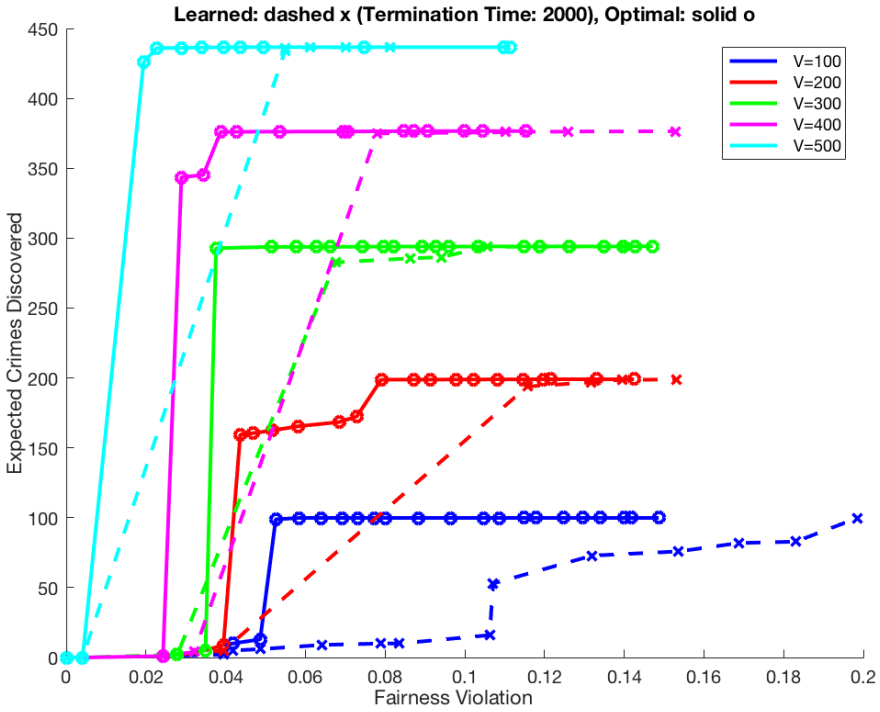


Figure 15: Pareto frontier of expected crimes discovered versus fairness violation.

uncensored observations. Censoring happens frequently because the number of police used in every case plotted is less than the daily average of 563.88 crimes across all the districts in the dataset — so it is unavoidable that in any allocation, there will be significant censoring in at least some districts.

Figure 15 shows that while the learning curves are dominated by the optimal curves, the performance of the learning algorithm approaches the performance of the offline optimal allocation as $\mathcal{V}$ increases. Again, this is because increasing $\mathcal{V}$ generally decreases the frequency of censoring. We study the $\mathcal{V} = 500$ regime in more detail, to explore the empirical rate of convergence. In Figure 16, we study the round by round performance of the allocation computed by Algorithm 2 in a single run with the choice of $\mathcal{V} = 500$ and $\alpha = 0.05$.

In Figure 16, the x axis labels progression of rounds of the algorithm. The y axis measures the fairness violation (left) and expected number of crimes discovered (right) of the alloca-

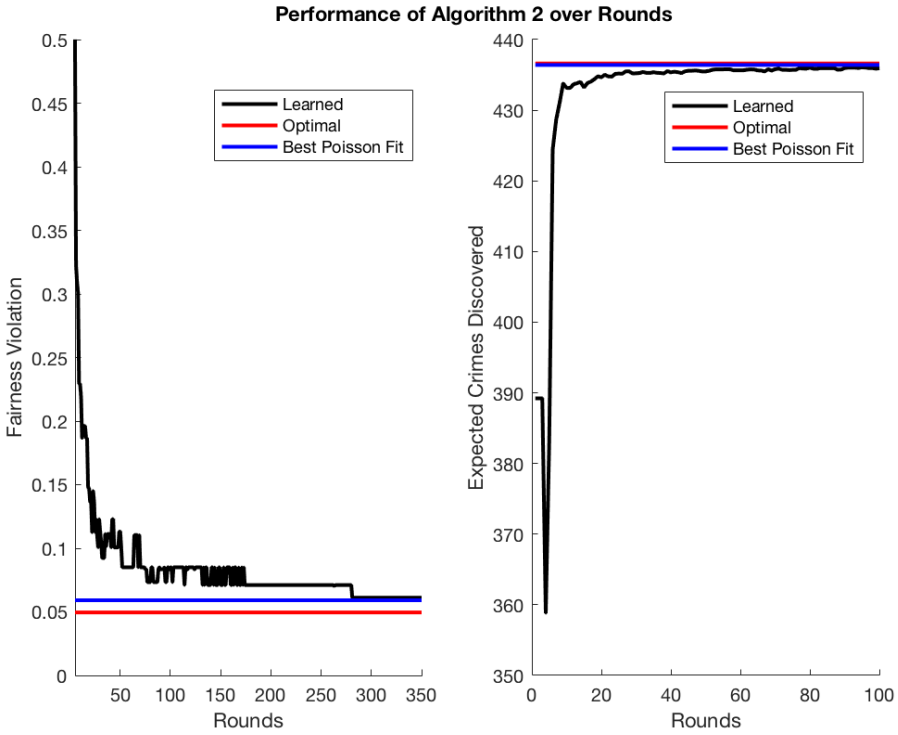


Figure 16: The per round expected number of crimes discovered and fairness violation of Algorithm 2. $\mathcal{V} = 500$ and $\alpha = 0.05$.

tion deployed by the algorithm, as measured with respect to the ground truth distributions. The black curves represent Algorithm 2. For comparison we also show the same quantities for an offline optimal fair allocation as computed with respect to the ground truth (red line), and an offline optimal fair allocation as computed with respect to the best Poisson fit to the ground truth (blue line). Note that in the limit, the allocations chosen by Algorithm 2 are guaranteed to converge to the blue baselines — but not the red baseline, because the algorithm is itself learning a Poisson approximation to the ground truth. The disparity between the red and blue lines quantifies the degradation in performance due to using Poisson approximations, rather than due to non-convergence of the learning process.

Figure 16 shows that Algorithm 2 converges to the Poisson approximation baseline well before the termination time of 2000, and substantially before the convergence bound guaranteed by our theory. Examining the estimated Poisson parameters used internally by

Algorithm 2 reveals that although the allocation has converged to an optimal fair allocation, the estimated parameters have not yet converged to the parameters of the best Poisson fit in any of the districts. In particular, Algorithm 2 systematically underestimates the parameters in all of the districts: the correlation coefficient between the true and estimated parameters is 0.9975.

We see also in Figure 16 that convergence to the optimum expected number of discovered crimes occurs more quickly than convergence to the target fairness violation level. This is also apparent in Figure 15 where the learning and optimal Pareto curves are generally similar in terms of the maximum number of crimes discovered, while the fairness violations are higher in the learning curves.

6.5. The Random Discovery Model

Finally, we consider the *random model* of discovery. In the random model, when v_i units are allocated to a group with c_i candidates, the number of discovered candidates is a random variable corresponding to the number of candidates that appear in a uniformly random sample of v_i individuals from a group of size m_i. Equivalently, when v_i units are allocated to a group of size m_i with c_i candidates, the number of candidates discovered by $\text{disc}(\cdot)$ is a random variable $\text{disc}(v_i, c_i) \triangleq o_i$, where o_i is drawn from the hypergeometric distribution with parameters m_i, c_i and v_i. Furthermore, the expected number of candidates discovered when allocating v_i units to group i is $\mathbb{E}[\text{disc}(v_i, c_i)] = v_i \mathbb{E}[c_i]/m_i$.

For simplicity, throughout this section, we assume $m_i \geq \mathcal{V}$ for all i. This assumption can be completely relaxed. Moreover, let $\mu_i = \mathbb{E}[c_i]/m_i$ denote the expected fraction of candidates in group i. Without loss of generality, for the rest of this section, we assume $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_{\mathcal{G}}$.

6.5.1. Optimal Allocation

In this section, we characterize optimal allocations. Note that the expected number of candidates discovered by the allocation choice $v_i \leq m_i$ in group i is simply $v_i\mu_i$. This suggests a simple algorithm to compute $\mathbf{w}^*$: allocating every unit of the resource to group 1. More generally, let $\mathcal{G}^* = \{i \mid \mu_i = \mu_1\}$ denote the subset of groups with the highest expected number of candidates. An allocation is optimal if and only if it only allocates *all* resources to groups in $\mathcal{G}^*$.

6.5.2. Properties of Fair Allocations

We next discuss the properties of fair allocations in the random discovery model. First, we point out that the discovery probability can be simplified as

$$f_i(v_i) = \mathbb{E}_{c_i \sim \mathcal{C}_i}\left[\frac{c_i v_i/m_i}{c_i}\right] = \frac{v_i}{m_i}.$$

So an allocation is α-fair in the random model if $|v_i/m_i - v_j/m_j| \leq \alpha$ for all groups i and j. Therefore, fair allocations (roughly) distribute resources in proportion to the size of the groups, essentially ignoring the candidate distributions within each group.

6.5.3. Price of Fairness

Recall that PoF quantifies the extent to which constraining the allocation to satisfy α-fairness degrades utility. While in Section 6.4 we study the PoF on the Philadelphia Crime Incidents dataset, we can define a worst-case variant as follows.

Definition 16. *Fix the random model of crime discovery and let $\alpha \in [0, 1]$. We define the PoF as*

$$PoF(\alpha) = \max_{\mathcal{C}} \frac{\chi(\boldsymbol{w}^*, \mathcal{C})}{\chi(\boldsymbol{w}^\alpha, \mathcal{C})}.$$

where $\mathcal{C}$ ranges over all possible candidate distributions.

We can fully characterize this worst-case PoF in the random discovery model.

Theorem 13. *The PoF in the random discovery model is*

$$PoF(\alpha) = \begin{cases} 1, & \frac{\mathcal{V}}{m_1} \leq \alpha, \\ \frac{M}{m_1 + \alpha(M - m_1)}, & \frac{\mathcal{V}}{m_1} > \alpha. \end{cases}$$

The PoF in the random model can be as high as M/m_1 in the worst case. If all groups are identically sized, this grows linearly with the number of groups.

6.6. Conclusion and Future Directions

Our presentation of allocative fairness provides a family of fairness definitions, modularly parameterized by a "discovery model". What counts as "fair" depends a great deal on the choice of discovery model, which makes explicit what would otherwise be unstated assumptions about the process of tasks like policing. The random and precision models of discovery studied in this paper represent two extreme points of a spectrum. In the predictive policing setting, the random model of discovery assumes that officers have no advantage over random guessing when stopping individuals for further inspection. The precision model assumes they can oracularly determine offenders, and stop only them. An interesting direction for future work is to study discovery models in between these two.

We have also made a number of simplifying assumptions e.g. we assumed the candidate distributions are *stationary* — fixed independently of the actions of the algorithm. Of course, the deployment of police officers can *change* crime distributions. Modeling this kind of dynamics, and designing learning algorithms that perform well in such dynamics would be interesting. Finally, we assumed that the same discovery model applies to all groups. One friction to fairness that one might reasonably conjecture is that the discovery model may differ between groups — being closer to the precision model for one group, and closer to the random model for another. We leave the study of these extensions to future work.

BIBLIOGRAPHY

[1] Philip Adler, Casey Falk, Sorelle Friedler, Gabriel Rybeck, Carlos Scheidegger, Brandon Smith, and Suresh Venkatasubramanian. Auditing black-box models for indirect influence. In *Proceedings of the 16th International Conference on Data Mining*, pages 1–10, 2016. 62, 68

[2] Alekh Agarwal, Peter Bartlett, and Max Dama. Optimal allocation strategies for the dark pool problem. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, pages 9–16, 2010. 81, 82

[3] Noga Alon. A note on network reliability. *Discrete Probability and Algorithms*, pages 11–14, 1995. 40

[4] Tansu Alpcan and Tamer Baar. *Network Security: A Decision and Game-Theoretic Approach*. Cambridge University Press, 1st edition, 2010. ISBN 0521119324, 9780521119320. 14

[5] Ross Anderson. *Security Engineering: A Guide to Building Dependable Distributed Systems*. Wiley Publishing, 2nd edition, 2008. ISBN 9780470068526. 14

[6] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. Machine bias. *Propublica*, 2016. 44, 62, 70

[7] James Aspnes, Kevin Chang, and Aleksandr Yampolskiy. Inoculation strategies for victims of viruses and the sum of squares partition problem. *Journal of Computer and System Sciences*, 72(6):1077–1093, 2006. 14

[8] Sepehr Assadi, Justin Hsu, and Shahin Jabbari. Online assignment of heterogeneous tasks in crowdsourcing markets. In *Proceedings of the Third AAAI Conference on Human Computation and Crowdsourcing*, pages 12–21, 2015. 6

[9] Venkatesh Bala and Sanjeev Goyal. A noncooperative model of network formation. *Econometrica*, 68(5):1181–1230, 2000. 1, 7, 8, 14, 16, 20, 32, 33, 34, 36, 38, 39, 41, 42

[10] Anna Barry-Jester, Ben Casselman, and Dana Goldstein. The new science of sentencing. *The Marshall Project*, August 8 2015. URL https://www.themarshallproject.org/2015/08/04/the-new-science-of-sentencing/. Retrieved 4/28/2016. 3, 44, 62

[11] Hamsa Bastani, Mohsen Bayati, and Khashayar Khosravi. Exploiting the natural exploration in contextual bandits. *CoRR*, abs/1704.09011, 2017. 82

[12] Richard Berk, Hoda Heidari, Shahin Jabbari, Matthew Joseph, Michael Kearns, Jamie Morgenstern, Seth Neel, and Aaron Roth. A convex framework for fair regression. In *4th Workshop on Fairness, Accountability, and Transparency in Machine Learning*, 2017. 6, 77

[13] Richard Berk, Hoda Heidari, Shahin Jabbari, Michael Kearns, and Aaron Roth. Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research*, 2018. 6, 77

[14] Sarah Bird, Solon Barocas, Kate Crawford, Fernando Diaz, and Hanna Wallach. Exploring or exploiting? social and ethical implications of autonomous experimentation in AI. 2016. 82

[15] Francis Bloch and Matthew Jackson. Definitions of equilibrium in network formation games. *International Journal of Game Theory*, 34(3):305–318, 2006. ISSN 0020-7276. 13

[16] Larry Blume, David Easley, Jon Kleinberg, Robert Kleinberg, and Éva Tardos. Network formation in the presence of contagious risk. In *Proceedings of the 12th ACM Conference on Electronic Commerce*, pages 1–10, 2011. 7, 8, 15, 38

[17] Ronen Brafman and Moshe Tennenholtz. R-MAX - A general polynomial time algorithm for near-optimal reinforcement learning. *Journal of Machine Learning Research*, 3:213–231, 2002. 55

[18] Nanette Byrnes. Artificial intolerance. *MIT Technology Review*, March 28 2016. URL https://www.technologyreview.com/s/600996/artificial-intolerance/. Retrieved 4/28/2016. 3, 44, 62

[19] Toon Calders and Sicco Verwer. Three naive Bayes approaches for discrimination-free classification. *Data Mining and Knowledge Discovery*, 21(2):277–292, 2010. 4, 62, 68

[20] Toon Calders, Asim Karim, Faisal Kamiran, Wasif Ali, and Xiangliang Zhang. Controlling attribute effect in linear regression. In *Proceedings of the 13th International Conference on Data Mining*, pages 71–80, 2013. 62, 69, 77

[21] Diego Cerdeiro, Marcin Dziubinski, and Sanjeev Goyal. Contagion risk and network design. *Working Paper*, 2014. 8, 14, 15

[22] Yu Chen, Shahin Jabbari, Michael Kearns, Sanjeev Khanna, and Jamie Morgenstern. Network formation under random attack and probabilistic spread. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 2019. 6

[23] Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, and Sergei Vassilvitskii. Fair clustering through fairlets. In *Proceedings of the 31th Annual Conference on Neural Information Processing Systems*, pages 5029–5037, 2017. 77

[24] Alexandra Chouldechova. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. In *3rd Workshop on Fairness, Accountability, and Transparency in Machine Learning*, 2017. 62, 64, 68, 75

[25] Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 797–806, 2017. 63, 70, 77

[26] William Cunningham. Optimal attack and reinforcement of a network. *Journal of ACM*, 32(3):549–561, 1985. ISSN 0004-5411. 14

[27] Kate Donahue and Jon Kleinberg. Fairness and utilization in allocating resources with uncertain demand. *CoRR*, abs/1906.09050, 2019. 93

[28] Jinshuo Dong, Hadi Elzayn, Shahin Jabbari, Michael Kearns, and Zachary Schutzman. Equilibrium characterization in data acquisition games. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 2019. 6

[29] Cynthia Dwork and Christina Ilvento. Fairness under composition. In *Proceedings of the 10th Innovations in Theoretical Computer Science*, 2019. 77

[30] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science*, pages 214–226, 2012. 3, 44, 46, 62, 69, 72, 77

[31] Hadi Elzayn, Shahin Jabbari, Christopher Jung, Michael Kearns, Seth Neel, Aaron Roth, and Zachary Schutzman. Fair algorithms for learning in allocation problems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 170–179, 2019. 6

[32] Danielle Ensign, Sorelle Friedler, Scott Neville, Carlos Scheidegger, and Suresh Venkatasubramanian. Runaway feedback loops in predictive policing. In *Conference on Fairness, Accountability and Transparency*, pages 160–171, 2018. 77, 78

[33] Danielle Ensign, Sorelle Frielder, Scott Neville, Carlos Scheidegger, and Suresh Venkatasubramanian. Decision making with limited feedback. In *Proceedings of the 29th Conference on Algorithmic Learning Theory*, pages 359–367, 2018. 77, 81

[34] Alex Fabrikant, Ankur Luthra, Elitza Maneva, Christos Papadimitriou, and Scott Shenker. On a network creation game. In *Proceedings of the 22nd ACM Symposium on Principles of Distributed Computing*, pages 347–351, 2003. 7, 8, 32

[35] Michael Feldman, Sorelle Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. Certifying and removing disparate impact. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 259–268, 2015. 44, 62, 68

[36] Benjamin Fish, Jeremy Kun, and Ádám Dániel Lelkes. A confidence-based approach for balancing fairness and accuracy. In *Proceedings of the 16th SIAM International Conference on Data Mining*, pages 144–152, 2016. 62, 63, 68, 69

[37] Sorelle Friedler, Carlos Scheidegger, and Suresh Venkatasubramanian. On the (im)possibility of fairness. *CoRR*, abs/1609.07236, 2016. 3, 47, 64, 68, 75

[38] Tobias Friedrich, Sven Ihde, Christoph Keßler, Pascal Lenzner, Stefan Neubert, and David Schumann. Efficient best response computation for strategic network formation under attack. In *Proceedings of the 10th International Symposium on Algorithmic Game Theory*, pages 199–211, 2017. 16, 37

[39] Kuzman Ganchev, Michael Kearns, Yuriy Nevmyvaka, and Jennifer Wortman Vaughan. Censored exploration and the dark pool problem. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence*, pages 185–194, 2009. 81, 85

[40] Sanjeev Goyal. *Connections: An Introduction to the Economics of Networks*. Princeton University Press, 2007. 15

[41] Sanjeev Goyal. Conflicts and networks. *The Oxford Handbook on the Economics of Networks.*, 2015. 14

[42] Sanjeev Goyal, Shahin Jabbari, Michael Kearns, Sanjeev Khanna, and Jamie Morgenstern. Strategic network formation with attack and immunization. In *Proceedings of 12th International Conference on Web and Internet Economics*, 2016. 6, 32, 33, 34, 37, 38, 39, 41, 42

[43] Assane Gueye, Jean Walrand, and Venkat Anantharam. A network topology design game: How to choose communication links in an adversarial environment. In *Proceedings of the 2nd International ICTS Conference on Game Theory for Networks*, 2011. 14

[44] Sara Hajian and Josep Domingo-Ferrer. A methodology for direct and indirect discrimination prevention in data mining. *IEEE Transactions on Knowledge and Data Engineering*, 25(7):1445–1459, 2013. 4, 46, 62, 68

[45] Moritz Hardt, Eric Price, and Nathan Srebro. Equality of opportunity in supervised learning. In *Proceedings of the 30th Annual Conference on Neural Information Processing Systems*, pages 3315–3323, 2016. 3, 4, 5, 44, 47, 62, 68, 69, 77, 78

[46] David Inouye, Eunho Yang, Genevera Allen, and Pradeep Ravikumar. A review of multivariate distributions for count data derived from the poisson distribution. *Wiley Interdisciplinary Reviews: Computational Statistics*, 9, 2017. 89

[47] Shahin Jabbari, Ryan Rogers, Aaron Roth, and Steven Wu. Learning from rational behavior: Predicting solutions to unknown linear programs. In *Advances in Neural Information Processing Systems 29*, pages 1570–1578, 2016. 6

[48] Shahin Jabbari, Matthew Joseph, Michael Kearns, Jamie Morgenstern, and Aaron Roth. Fairness in reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning*, pages 1617–1626, 2017. 6, 77

[49] Matthew Jackson. *Social and Economic Networks*. Princeton University Press, 2008. 31

[50] Zubin Jelveh and Michael Luca. Towards diagnosing accuracy loss in discrimination-aware classification: An application to predictive policing. In *2nd Workshop on Fairness, Accountability, and Transparency in Machine Learning*, 2015. 63

[51] Kory Johnson, Dean Foster, and Robert Stine. Impartial predictive modeling: Ensuring fairness in arbitrary models. *CoRR*, abs/1608.00528, 2016. 69

[52] Matthew Joseph, Michael Kearns, Jamie Morgenstern, and Aaron Roth. Fairness in learning: Classic and contextual bandits. In *Proceedings of the 30th Annual Conference on Neural Information Processing Systems*, pages 325–333, 2016. 3, 4, 44, 45, 47, 50, 72, 77

[53] Faisal Kamiran and Toon Calders. Data preprocessing techniques for classification without discrimination. *Knowledge and Information Systems*, 33(1):1–33, 2011. 4, 62, 68, 69

[54] Faisal Kamiran, Toon Calders, and Mykola Pechenizkiy. Discrimination aware decision tree learning. In *Proceedings of the 10th IEEE International Conference on Data Mining*, pages 869–874, 2010.

[55] Faisal Kamiran, Asim Karim, and Xiangliang Zhang. Decision theory for discrimination-aware classification. In *Proceedings of the 12th IEEE International Conference on Data Mining*, pages 924–929, 2012. 4, 46

[56] Toshihiro Kamishima, Shotaro Akaho, Hideki Asoh, and Jun Sakuma. Fairness-aware classifier with prejudice remover regularizer. In *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 35–50, 2012. 46, 62, 68, 69

[57] Sampath Kannan, Jamie Morgenstern, Aaron Roth, Bo Waggoner, and Zhiwei Steven Wu. A smoothed analysis of the greedy algorithm for the linear contextual bandit problem. In *Proceedings of the 32nd Annual Conference on Neural Information Processing Systems*, 2018. 82

[58] Michael Kearns and Luis Ortiz. Algorithms for interdependent security games. In *Proceedings of the Advances in Neural Information Processing Systems 16*, pages 561–568, 2003. 14

[59] Michael Kearns and Satinder Singh. Near-optimal reinforcement learning in polynomial time. *Machine Learning*, 49(2-3):209–232, 2002. 46, 49, 58

[60] Michael Kearns, Yishay Mansour, and Andrew Ng. Approximate planning in large POMDPs via reusable trajectories. In *Proceedings of the 13th Annual Conference on Neural Information Processing Systems*, pages 1001–1007, 2000. 58

[61] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 137–146, 2003. 2, 30, 32, 35, 37

[62] Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. Inherent trade-offs in the fair determination of risk scores. In *Proceedings of the 7th Conference on Innovations in Theoretical Computer Science*, 2017. 47, 64, 68, 75

[63] Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. Inherent trade-offs in the fair determination of risk scores. In *Proceedings of the 8th Conference on Innovations in Theoretical Computer Science*, pages 43:1–43:23, 2017. 3, 77

[64] Lasse Kliemann. The price of anarchy for network formation in an adversary model. *Games*, 2(3):302–332, 2011. 15, 38

[65] Lasse Kliemann, Elmira Shirazi Sheykhdarabadi, and Anand Srivastav. Swap equilibria under link and vertex destruction. *Games*, 8(1):14, 2017. 38

[66] Ron Kohavi. Scaling up the accuracy of naive Bayes classifiers: A decision-tree hybrid. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, pages 202–207, 1996. 70

[67] Aron Laszka, Dávid Szeszlér, and Levente Buttyán. Linear loss function for the network blocking game: An efficient model for measuring network robustness and link criticality. In *Proceedings of the 3rd International Conference on Decision and Game Theory for Security*, pages 152–170, 2012. 14

[68] Pascal Lenzner. Greedy selfish network creation. In *Proceedings of the 8th International Workshop on Internet and Network Economics*, pages 142–155, 2012. 13

[69] Moshe Lichman. UCI machine learning repository, 2013. URL http://archive.ics.uci.edu/ml. 70

[70] Shiau Hong Lim, Huan Xu, and Shie Mannor. Reinforcement learning in robust markov decision processes. In *Proceedings of the 27th Annual Conference on Neural Information Processing Systems*, pages 701–709, 2013. 46

[71] Lydia Liu, Sarah Dean, Esther Rolf, Max Simchowitz, and Moritz Hardt. Delayed impact of fair machine learning. In *Proceedings of the 35th International Conference on Machine Learning*, pages 3156–3164, 2018. 77

[72] Kristian Lum and William Isaac. To predict and serve? *Significance*, pages 14–18, October 2016. 77

[73] Binh Thanh Luong, Salvatore Ruggieri, and Franco Turini. k-NN as an implementation of situation testing for discrimination discovery and prevention. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 502–510. ACM, 2011. 46, 62, 68

[74] Wolfgang Mader. Über die maximalzahl kreuzungsfreier h-wege. *Archiv der Mathematik (Basel)*, 31:387–402, 1978. 19

[75] Shie Mannor, Ofir Mebel, and Huan Xu. Lightning does not strike twice: Robust MDPs with coupled uncertainty. In *Proceedings of the 29th International Conference on Machine Learning*, 2012. 46

[76] Clair Miller. Can an algorithm hire better than a human? *The New York Times*, June 25 2015. URL http://www.nytimes.com/2015/06/26/upshot/can-an-algorithm-hire-better-than-a-human.html/. Retrieved 4/28/2016. 3, 44, 62

[77] Jun Morimoto and Kenji Doya. Robust reinforcement learning. *Neural computation*, 17(2):335–359, 2005. 46

[78] Dino Pedreshi, Salvatore Ruggieri, and Franco Turini. Discrimination-aware data mining. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 560–568. ACM, 2008. 3, 46, 62, 68

[79] Ariel Procaccia. Cake cutting: Not just child's play. *Communications of the ACM*, 56 (7):78–87, 2013. 82

[80] Manish Raghavan, Aleksandrs Slivkins, Jennifer Wortman Vaughan, and Zhiwei Steven Wu. The externalities of exploration and how data diversity helps exploitation. In *Proceedings of the 31st Conference On Learning Theory*, pages 1724–1738, 2018. 82

[81] Sankardas Roy, Charles Ellis, Sajjan Shiva, Dipankar Dasgupta, Vivek Shandilya, and Qishi Wu. A survey of game theory as applied to network security. In *Proceedings of the 43rd Hawaii International Conference on System Sciences*, pages 1–10, 2010. ISBN 978-0-7695-3869-3. 14

[82] Cynthia Rudin. Predictive policing using machine learning to detect patterns of crime. *Wired Magazine*, August 2013. URL http://www.wired.com/insights/2013/08/predictive-policing-using-machine-learning-to-detect-patterns-of-crime/. Retrieved 4/28/2016. 44, 62

[83] Satinder Singh. Personal Communication, June 2016. 50

[84] Richard Sutton and Andrew Barto. *Introduction to Reinforcement Learning*. MIT Press, Cambridge, MA, USA, 1st edition, 1998. 48

[85] Latanya Sweeney. Discrimination in online ad delivery. *Communications of the ACM*, 56(5):44–54, 2013. 3, 44, 62

[86] Istv'an Szita and Csaba Szepesvári. Model-based reinforcement learning with nearly tight exploration complexity bounds. In *Proceedings of the 27th International Conference on Machine Learning*, pages 1031–1038, 2010. 46

[87] Chi Wang, Wei Chen, and Yajun Wang. Scalable influence maximization for independent cascade model in large-scale social networks. *Data Mining and Knowledge Discovery*, 25(3):545–576, 2012. 30, 33, 36

[88] Blake Woodworth, Suriya Gunasekar, Mesrob Ohannessian, and Nathan Srebro. Learning non-discriminatory predictors. In *Proceedings of the 30th Conference on Learning Theory*, 2017. 4, 47, 68, 77

[89] I-Cheng Yeh and Che-hui Lien. The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications*, 36(2):2473–2480, 2009. 70

[90] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez-Rodriguez, and Krishna P. Gummadi. Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1171–1180, 2017. 3, 47, 63, 68, 69, 77

[91] Richard Zemel, Yu Wu, Kevin Swersky, Toniann Pitassi, and Cynthia Dwork. Learning fair representations. In *Proceedings of the 30th International Conference on Machine Learning*, pages 325–333, 2013. 46, 69, 77