DESIGNING ALGORITHMS FOR SOCIAL GOOD

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Rediet Tesfaye Abebe

December 2019

DESIGNING ALGORITHMS FOR SOCIAL GOOD

Rediet Tesfaye Abebe, Ph.D.

Cornell University 2019

Algorithmic and artificial intelligence techniques show immense potential to deepen our understanding of socioeconomic inequality and inform interventions designed to improve access to opportunity. Interventions aimed at historically underserved communities are made particularly challenging by the fact that disadvantage and inequality are multifaceted, notoriously difficult to measure, and reinforced by feedback loops in underlying structures. While great strides have been made in these areas – from assigning seats in public schools to poverty mapping – there remain many domains with major opportunities for further contributions and the prospect that we may be able to develop unified frameworks for applying computational insights to improve societal welfare.

In this thesis, we develop algorithmic and computational techniques to address these issues through two types of interventions: one in the form of allocating scarce societal resources and the other in the form of improving access to information. We examine the ways in which techniques from algorithms, discrete optimization, mechanism design, and network and computational sciences can combat different forms of disadvantage, including susceptibility to income shocks, social segregation, and disparities in access to health information. We highlight opportunities for computing to play a role in fundamental social change. We close with a discussion on open questions in an emerging research area – *Mechanism Design for Social Good* (MD4SG) – around the use of algorithms, optimization, and mechanism design to address.

BIOGRAPHICAL SKETCH

Rediet Abebe is a computer science researcher, broadly working in the areas of algorithms and artificial intelligence with a focus on their applications to equity and social good concerns. As part of this research agenda, she co-founded the *Mechanism Design for Social Good* (MD4SG) research initiative as a Ph.D. student at Cornell University. After graduating from Cornell, she will continue her tenure as a Junior Fellow at Harvard University, Society of Fellows.

Abebe is an alumna of Harvard University (M.S. in applied mathematics, 2015), University of Cambridge (M.A. in mathematics, 2014), Harvard College (B.A. in mathematics, 2013), as well as the International Community School of Addis Ababa and Nazareth School. To foster inclusion and representation in her field, Abebe co-founded the *Black in AI* network as a graduate student. Her work is deeply informed by her upbringing in Addis Ababa, Ethiopia.

For my mother and the motherland.

ACKNOWLEDGEMENTS

I knew I wanted to be an academic at a young age and I did not think that it would be possible for me to become any more excited about the prospect of being a researcher, mentor, and teacher than when I started graduate school. I feel immense gratitude to Jon Kleinberg for proving me wrong and showing me new ways to be excited about this career from my very first visit to Cornell. Thank you for your guidance, mentorship, and friendship, and for teaching me the value of a well-placed em dash – I could not be more grateful.

I have been lucky to have a secondary advisor in David Parkes who, in addition to getting me started on this research path, helped me as I transitioned from mathematics to computer science, from Harvard to Cornell then back again. Your generous support and teaching through the years has helped me achieve this milestone; thank you very much.

I have learned an incredible amount from Bobby Kleinberg, who, in addition to being the ultimate test of my most challenging math problems, has been an unrelentless cheerleader for me over the years. Thank you for everything you have taught me and for believing in me since the first moment we met.

I am deeply grateful, also, for Michael Macy, for welcoming me into your lab, always asking questions that I would not have thought of, and the most timely boosts in confidence. I am also thankful to my committee member, Kilian Weinberger, for your guidance and support, despite my refusal to do deep learning. Maybe someday, Kilian.

I have benefited greatly from being a member of the theory of computation group, the AI, Policy, and Practice group, and the Center for the Study of Inequality. I would especially like to thank Karen Levy and Kim Weeden for being incredible collaborators, mentors, and friends. Many of the ideas presented in this thesis have their origin in presentations I attended in one of these initiatives or discussions in Kim Weeden's social stratification class. Thank you for everything.

I have also benefited from working with each of my collaborators on the work presented in this thesis: Lada Adamic, Solon Barocas, Richard Cole, Vasilis Gkatzelis, Jason Hartline, Shawndra Hill, Jon Kleinberg, Karen Levy, Nicole Immorlica, Brendan Lucier, H. Andrew Schwartz, Peter M. Small, Manish Raghavan, David Robinson, Jennifer Wortman Vaughan, Matt Weinberg. Thank you all. Two of the chapters in this thesis started as a result of research internships at Microsoft Research, where I was mentored by Nicole Immorlica, Brendan Lucier, and Jennifer Wortman Vaughan. Thank you for providing these opportunities for me to grow as a researcher.

My research has been enriched by the many people I have had the pleasure of working with through MD4SG. Many thanks to Kira Goldner for your incredible insight and vision in building the foundation over the years, Irene Lo for taking the group to a whole new level, and Ana-Andreea Stoica for effortlessly jumping in and showing us creative ways to expand our purpose. I have learned a lot and been inspired by each of you. I am grateful for Francisco Marmolejo, Bryan Wilder, Zöe Hitzig, Sara Kingsley, Faidra Monachou, Moses Namara, Manish Raghavan, Swathi, Sadagopan, Sam Taggart, and many others for leading groups within MD4SG and additionally my collaborators Kehinde Aruleba, Basiliyos Betru, Lenore Cowen, Rachel Cummings, Sanmay Das, Ezra Gross, Cynthia Habomania, Sera Linardi, Dina Machuve, Alvaro Monge, Daniel Nkemelu, Ezinne Nwankwo, George Obaido, Sekou Remy, Mwiza Simbeye, and many others.

I would also like to thank many members of the Cornell CIS community.

Thank you to David Bindel, Éva Tardos, and Hakim Weatherspoon for your academic and personal mentorship of me and other students, and especially those of us from underrepresented groups. Thank you to my collaborator and friend Austin Benson for your inspiring creativity and unending support. Thank you to Rebecca Stewart, Yl Guanchez, Vanessa Maley, and Coralia Torees for all of your support and patience with the unending stream of tasks for Black in AI and MD4SG, among many other things. I am very grateful for you all. And many thanks to numerous other Cornell CIS faculty, staff, and students, who made my time at Cornell all the more better.

Thank you to everyone on Cornell's West Campus, and especially to Erica Ostermann and Julia Thom-Levy for the community you created in Bethe House, which was my home for the majority of my graduate studies at Cornell.

Part of the work presented in this thesis was completed during my first semester at the Harvard Society of Fellows. I would like to thank Kelly Katz, Ana Novak, and Diana Morse for your support during my transition into the Society, Sarah Derbew and Mireille Kamariza for all that you do to create community, and the Senior Fellows for this wonderful gift of time.

I credit much of the encouragement and sense of belonging I have felt as a graduate student to the Black in AI community, which continues to be one of the greatest sources of joy for me. Many thanks to Esube Bekele, Sarah Brown, Moustapha Cisse, Hassan Kane, Sanmi Koyejo, Daniel Nkemelu, Ezinne Nwankwo, Lyne Tchapmi, Judy Wawira, and especially to Timnit Gebru and Devin Guillory for all your work building this community and your support and friendship.

My academic career and transition into computer science was instigated and facilitated by summers I spent working with László Babai and Peter May, who continue to be two of my most-trusted mentors and friends. Thank you both for taking me under your wing and encouraging me to be a researcher, despite never having a formal obligation in my training.

I am also grateful for Imre Leader and Felix Fischer for supervising my research and studies during my time at the University of Cambridge, to Barbara Grosz, Radhika Nagpal, and David Parkes for getting me excited about AI during my year at Harvard SEAS, and to Michael Hopkins, Richard Stanley, and Victor Reiner for getting me started on research as an undergraduate student.

Growing up I benefited from being a member of many communities in Ethiopia, including the International Community School of Addis Ababa and Nazareth School, of which I am a proud alumna.

And, last, I am grateful for my families, God-given, assigned, and chosen: thank you to the Neale Parises: to Carly, Jasmine, Tallulah, and Leo for welcoming me as one of the sisters from the first day and to Bridget and Matteo for opening up your home and your heart – your friendship means the world to me. To Fiona Wood and Diana Cai: thank you for being the kindest, wisest, and most generous friends I could have asked for. Thank you to David Miller for your boundless support, always knowing when to affirm and when to challenge me, and, of course, thank you for typing. Thank you to Ililan and Tsega Abebe for being my first and dearest friends. Thank you to Tesfaye Abebe for being the first to see a scientist in me. And, most of all, thank you to Assegedech Ketema Abebe for always being my greatest joy of inspiration and courage in this and everything.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Part I

Introduction and Background

CHAPTER 1

INTRODUCTION

Algorithms and artificial intelligence have shown promise to deepen our understanding of disadvantage and socioeconomic inequality. None of the dimensions of inequality – whether economic, social, and cultural – alone conveys the whole picture; many, furthermore, are complex and difficult to measure (Grusky, 2018; Grusky and Weeden, 2016; Tumin et al., 2012). These challenges are further exacerbated by feedback loops in the underlying socioeconomic structures. There is thus an opportunity for algorithmic and computational insights to inform interventions that can help address the needs of individuals and communities

Thus far, computational applications have ranged from mechanism design for kidney exchange (Roth et al., 2005, 2004), to machine learning for poverty mapping (Jean et al., 2016; Xie et al., 2016), and developing technology for under-resourced communities (Patel et al., 2010). While great strides have been made in many of these areas, there remain numerous domains with major opportunities for exploration and the potential for developing unified frameworks to identify opportunities for computational insights to help improve societal welfare.

In this thesis, we present findings at this research interface of algorithms, artificial intelligence, and social sciences. Specifically, we develop and employ algorithmic and computational techniques to address the above issues through two types of interventions discussed in Parts II and III:

Part II Designing algorithmic, optimization, and mechanism design-based approaches to improve allocation of societal resources, especially in settings in which such resources may be scarce or there is heterogeneity in need or preference for these resources, and

Part III Leveraging computational and network-based insights to measure and improve access to information for disadvantaged groups.

In doing so, we make contributions towards fundamental algorithmic questions – such as the *house allocation* problem in matching and mechanism design – identify new directions for contributions, including at the interface of public finance and optimization, and conduct the, to our knowledge, first-ever published academic study using large web and social media data to study health across all 54 nations in Africa. We highlight technical challenges that arise in these settings, ranging from lack of access to ground-truth data and adequate measurements of need or welfare to impossibility results around allocation mechanisms satisfying certain fairness and efficiency criteria. Tackling these problems, we examine the ways in which our techniques can combat different forms of disadvantage. Some examples which will be presented in this thesis include susceptibility to income shocks, social segregation, and disparities in access to health information.

We discuss these results in the context of the relevant domains. In doing so, we draw on insights from development, economics, global health, public policy, science and technology studies (STS), and sociology. In addition to these above results, in **Part IV**, we further provide a framework for roles that computing – specifically algorithm design, machine learning, and mechanism design – can play in fundamental social change and discuss ethical challenges.

Blending these takeaways, we close with a discussion of an emerging research area, *Mechanism Design for Social Good* (MD4SG), which uses techniques from algorithms, optimization, and mechanism design, along with insights from the social sciences and humanistic studies, to help improve access to opportunity for historically under-served and disadvantaged communities (Abebe and Goldner, 2018a,b). We provide a menu of open directions from the MD4SG research community in **Part V**.

The published version of the work presented in this thesis can be found in Abebe et al. (2018, 2020a, 2019a); Abebe and Goldner (2018a); Abebe et al. (2019b,c, 2020b). We begin by summarizing the main contributions of each of the subsequent chapters:

1.1 Summary of Results

1.1.1 Societal Resource Allocations

In Part II, we study allocation problems that arise in many public sector settings. We design and investigate optimization and mechanism design-based schemes that can inform allocation of societal resources.

Optimization-Based Subsidy Allocations. In Chapter 4, we tackle the problem of the multi-faceted nature of economic welfare. Assistance programs, such as those targeting housing instability, can face challenges when they rely on simpler measures of welfare such as income or wealth, since these fail to capture to full complexity of families' state. Informed by empirical work in public finance and sociology, we consider one important dimension – *susceptibility to income shocks*. We introduce a model of welfare that incorporates income, wealth, and income shocks. We analyze this model to show that it can vary, at times substantially, from measures of welfare that use only income and/or wealth.

We then study the algorithmic problem of optimally allocating subsidies in the presence of income shocks. We consider two well-studied objectives: the first aims to minimize the expected number of agents that fall below a given welfare threshold (a *min-sum* objective) and the second aims to minimize the likelihood that the most vulnerable agent falls below this threshold (a *min-max* objective). We consider subsidies to agents' income and wealth and give an efficient optimal allocation mechanism for the min-max objective under a general setting. Likewise, we show that we can give an optimal allocation mechanism for the min-sum objective under natural assumptions on the agents' wealth and shock distributions. Namely, we consider cases in which the agents have zero wealth or those in which the shocks are drawn from an exponential distribution, and give polynomial-time algorithms for allocating subsidies for these settings. Then, for the general case, we give a *fully polynomial-time approximation scheme* (FPTAS) to optimize for the min-sum objective. We provide a discussion of open questions related to these problems and societal implications of such optimization-based allocation schemes. In Part V, we discuss further open directions in settings where measurements of welfare are challenging and resources may be scarce.

Mechanism Design for Allocations. The above problem considers allocation of a homogeneous resource – such as subsidies – in a setting where the agents' income, wealth, and shock parameters are known. In certain allocation problems, however, different agents may have different preferences for the resources being allocated. Examples include allocation of seats in public schools or public housing (Abdulkadiroğlu and Sönmez, 2003; Arnosti and Shi, 2018; Thakral, 2016; Waldinger, 2017). Here, agents may have preferences by, for instance, location of school, programs offered at schools, housing types, or amenities provided. In these settings, we may need to solicit the agents' preferences, but the agents may strategically manipulate their preferences to improve their allocations. This presents a key challenge in *mechanism design*, a field of economics and computation that concerns itself with the design of algorithms where inputs, such as preferences, may be coming from strategic agents (Hartline, 2013; Nisan and Ronen, 2001).

A fundamental problem in mechanism design is what is known as the *house allocation* problem where a set of n agents need to be matched to a set of n heterogeneous, indivisible items (Hylland and Zeckhauser, 1979). This problem appears in many public sector settings where families may need to be matched one-to-one to a set of items, such as seats in public schools, houses or housing resources, and other public goods. In Chapter 5, we introduce a mechanism that truthfully elicits *cardinal* preferences from the agents and provides stronger performance guarantees than several other mechanisms in this setting. Specifically, we assume that each agent i has a value of $v_{i,j}$ for each item j, and these values are private information that the agents may misreport if doing so leads to a preferred outcome. Ensuring that the agents have no incentive to misreport requires a careful design of the matching mechanism, and mechanisms proposed in the literature mitigate this issue by eliciting only the *ordinal* preferences of the agents, i.e., their ranking of the items from most to least preferred. However, the efficiency guarantees of these mechanisms are based only on weak measures that are oblivious to the underlying values.

Our mechanism truthfully elicits the full cardinal preferences of the agents, i.e., all of the $v_{i,j}$ values. We evaluate the performance of this mechanism using the much more demanding Nash bargaining solution as a benchmark, and we prove that our mechanism significantly outperforms all ordinal mechanisms (even non-truthful ones). To prove our approximation bounds, we also study the population monotonicity of the Nash bargaining solution in the context of matching markets, providing both upper and lower bounds which are of independent interest. We close with a discussion of further open questions in mechanism design and in Part V, we provide further matching-based questions that can improve access to resources or information.

1.1.2 Improving Access to Information

Access to information is a major global concern. Focusing on health, for instance, it is known that *access to quality health information* can play a crucial role in mitigating the burden of diseases, reducing transmissions, and managing infections. There are two key challenges in improving access to health information. The first is finding high-quality data on the health information needs of key populations; the second is, upon identifying these needs, targeting education and information at these populations. In Part III, we tackle each of these challenges.

Measuring Access to Information. In Chapter 7, we turn to the problem of identifying health information needs of individuals in data-sparse regions, focusing on the African continent as a case study. The lack of comprehensive, high-quality health data in Africa creates a roadblock for combating disease. Without understanding people's everyday concerns, health organizations and policymakers are less able to effectively target education and programming efforts. In this chapter, we propose a bottom-up approach that uses search data to gain insight into health information needs of individuals in Africa. We analyze Bing searches related to HIV/AIDS, malaria, and tuberculosis from all 54 African nations. For each disease, we automatically derive a set of common topics, revealing a widespread interest in various types of information, including disease symptoms, drugs, concerns about breastfeeding, as well as stigma, beliefs in natural cures, and other topics that may be difficult to uncover through traditional surveys.

We expose the different patterns that emerge in health information needs by demographic groups (age and gender) and country. Using finer-grained data, we also uncover *discrepancies in the quality of content* returned by the search engines to users by topic and highlight differences in user behavior and satisfaction. Combined, our results suggest that search data can help illuminate health information needs in Africa and inform discussions on health policy and targeted education efforts, both on- and off-line. This is the first work to identify specific health information needs in all 54 nations in Africa using large-scale observational data from the web. In Part V, we highlight further opportunities to improve measurements of access to information for marginalized groups and where such data-sources can help address data inequalities.

Network-Based Interventions. Networks have long been understood to play a key role in the *diffusion of information* across populations, playing a significant role in the welfare of all communities (Easley and Kleinberg, 2010; Rogers, 1995). A popular phenomenon in social networks is word-of-mouth recommendations, in which information spreads from one person to another, potentially spreading across the full network (Domingos and Richardson, 2001; Kempe et al., 2003; Richardson and Domingos, 2002). *Influence maximization*, the process of selecting the most influential agents on a network in order to propagate information across the full population, is one popular strategy which has been employed in a variety of contexts including in health (Wakefield et al., 2003; Yadav et al., 2016, 2017, 2018).

In traditional models for word-of-mouth recommendations, the objective function has generally been based on reaching as many people as possible. However, a number of studies have shown that the indiscriminate spread of an innovation by word-of-mouth can result in *overexposure*, reaching people who evaluate it negatively. For instance, empirical work has demonstrated that targeting campaigns at populations that are not receptive to it (as with anti-smoking or anti-drug campaigns such as DARE) can, in fact, increase the behavior the campaigns are designed to discourage (such as tobacco and alcohol use) (Hamilton, 1997). We therefore ask, how should one make use of social influence when there is a risk of overexposure? In this Chapter 8, we develop and analyze a theoretical model for this process; we show how this model captures a number of qualitative phenomena associated with overexposure, and for the main formulation of our model, we provide a polynomial-time algorithm to find the optimal strategy. We also present simulations of the model on real network topologies, quantifying the extent to which our optimal strategies outperform natural baselines.

Diffusion of information on social networks is hampered by certain social processes, chief among which being *homophily* – the tendency for individuals to form social ties to others who are similar to themselves. Homophily is one of the most robust sociological principles, leading to patterns of linkage in social networks that segregate people along many different demographic dimensions (McPherson et al., 2001). This phenomenon, in turn, can result in inequalities in access to information and opportunities by members of different demographic groups (Avin et al., 2015; Calvo-Armengol et al., 2009; Calvo-Armengol and Jackson, 2004; Dasaratha, 2017; DiMaggio and Garip, 2011; Easley and Kleinberg, 2010; Stoica et al., 2018). As we consider potential interventions that might alleviate the effects of network segregation, we face the challenge that homophily constitutes a pervasive and organic force that is difficult to push back against. The design of interventions can therefore benefit from identifying counterbalancing natural processes in the network that might be harnessed to work in opposition to segregation.

In Chapter 9, we show that, contrary to popular belief, triadic closure may be one such process. *Triadic closure* — the process by which individuals are more likely to form social ties with other individuals with whom they already share ties – has long been believed to lead to further segregation in networks where agents show homophily when they form links (Granovetter, 1977; Kossinets and Watts, 2006; Rapoport, 1953; Stadtfeld, 2015; Tóth et al., 2019). The idea here is that triadic closure gives further exposure to like-type individuals on a network as any existing links are more likely to be between individuals that share a type. In this chapter, we demonstrate the counterintuitive power for triadic closure to reduce network segregation through analysis of several fundamental network models. The models that we consider here exhibit an interplay between triadic closure and homophily and are well-studied models in mathematics, computer science, and economics.

We leverage insights around triadic closure to discuss interventions based on link recommendation that can help reduce network segregation. We show that small nudges in the early stages of a network formation process can help move the network into a more overall integrated state. In Part V, we present further open directions related to networks, inequality, and access to information in settings inspired by education, health, and labor markets.

1.1.3 Roles for Computing in Social Change

A growing body of computing research, including work presented in this thesis, aims to respond to social problems. But some scholars have argued that such work may be counterproductive when it treats problematic features of the status quo as fixed, and may fail to address deeper patterns of racial, social and economic oppression. While acknowledging these critiques, we posit in Part IV that computational research has valuable roles to play in addressing social problems — roles whose utility can be recognized even from a perspective that aspires toward fundamental social change. We articulate four such roles, through an analysis that considers the opportunities as well as the significant risks inherent in such work. Computing research can serve as a *diagnostic*, helping us to understand and measure social problems with precision and clarity. As a *formalizer*, computing shapes how social problems are explicitly defined — changing how the problems, and possible responses to them, are understood. Computing serves as *rebuttal* when it illuminates the boundaries of what is possible through technical means. And computing acts as *synecdoche* when it makes long-standing social problems newly salient in the public eye. We offer these paths forward as modalities that leverage the particular strengths of computational work in the service of social change, without over-claiming computing's capacity to solve social problems on its own.

1.1.4 Open Directions

We close with a discussion of an emerging research area– *Mechanism Design for Social Good* (MD4SG) – which shares the research mission outlined in this thesis. We provide open research directions covering domains ranging from allocating housing and homelessness resources to improving access to opportunities in developing nations and mitigating inequality in on- and off-line labor markets. These research directions build on the work presented in this thesis as well as presentations and discussions in the MD4SG research community.[1] We present these by domain, providing further examples of opportunities for computing to play a role as an ally for broad social change. These directions are not intended to be exclusive or exhaustive, but rather to provide a sample of potential avenues for further exploration.

[1]The directions provided here are informed by articles, notes, and discussions within the MD4SG community (Abebe et al., b; Abebe and Goldner, 2018a,b).

CHAPTER 2

BACKGROUND AND RELATED WORK

2.1 Societal Resource Allocations

2.1.1 Optimization-Based Subsidy Allocations

The multi-faceted nature of economic welfare creates challenges for measuring and analyzing it in a tractable manner (Atkinson, 2015; Grusky, 2018; Grusky and Ku, 2008; Grusky and Weeden, 2016; Piketty, 2015; Sen, 1976; Sen et al., 1997). A key point in this area has been to introduce measurements that account for the complexity while preserving enough simplicity to inform policy and interventions aimed at mitigating poverty and economic inequality. In Chapter 4, we study one dimension in this space — the phenomenon of *income shocks* — whose impact on economic welfare has been extensively documented through a long line of empirical work.

Income shocks have been shown to have complex interactions with many factors and have long-term and severe consequences for families including eviction, significant drop in income, job loss, and poor health outcomes (Atake, 2018; Björkman-Nyqvist, 2013; Desmond, 2012, 2015, 2016; Desmond and Gershenson, 2016; Desmond and Shollenberger, 2015; Dinkelman et al., 2008; Giesbert and Schindler, 2012; Jarosch, 2015; Kijima et al., 2006; Kochar, 1995, 1999; Mello, 2018; Morduch, 1994; O'Flaherty, 2009; Shapiro et al., 2004). For a discussion on the role of income shocks on the well-being of families in New York City, see reports by the Poverty Tracker (Wimer et al., 2014a, 2016a, 2014b, 2016b; Lens et al., 2016).

Despite these observations, shocks do not play a correspondingly central role in many social assistance programs. In particular, in this chapter, we are concerned with the problem of informing allocation of subsidies in the presence of income shocks. There is a long line of work in the study of household consumptions and public economics that considers agent-level behavior in response to subsidies and more broadly understanding consumption dynamics (Aiyagari, 1994; Golosov et al., 2006; Kocherlakota, 2004) as well as optimal taxation theory (Eaton and Rosen, 1980a,b,c; Farhi and Werning, 2013; Saez and Stantcheva, 2018; Varian, 1980). The work presented in this chapter is informed by this above area. We however take an algorithmic approach and study the inherent stochastic optimization problems at the heart of subsidy allocation with shocks.

Our model is based on the framework of *ruin probabilities* from the literature on stochastic risk modeling (Asmussen and Albrecher, 2010). This work also has close connections with stochastic control theory. There has been work related to investment in risky assets by insurance companies seeking to minimize the probability of ruin (Azcue and Muler, 2009; Hipp and Plum, 2000; Schmidli et al., 2002), but to our knowledge there has not been work looking at our problem of allocating a fixed budget over multiple agents in order to change premium flow or initial wealth.

The use of computational techniques to fairly and efficiently allocate societal resources has a long history in the economics and computation literature. The problem we address in studying subsidy allocations is loosely connected to the fair division literature. For a survey on fair allocation, see Brams and Taylor

(1996); Moulin (2003); Procaccia (2013); Procaccia and Moulin (2016); Robertson and Webb (1998); Young (1995). Departing from this work, the item we consider here is homogeneous. Moreover, rather than continuously improving all or a subset of the agents' welfare, we are concerned with ensuring that agents do not fall below a certain threshold, but we cannot guarantee that they do not.

Our contribution can be viewed as belonging to an emerging style of work that uses computational and optimization-based methods to inform assistance programs aimed at improving access to opportunity for vulnerable populations (Abebe and Goldner, 2018a,b). One recent instance studies allocating interventions for homelessness services using a mixture of counterfactual reasoning and mechanism design (Kube et al., 2018); another studies optimal allocation of financial aid in US colleges based on students' parental income (Findeisen and Sachs, 2016). Relative to these recent papers, our work is the first that we are aware of to take a computational approach to assistance in a setting where the underlying optimization problem exhibits the type of rich stochastic dynamics over time that characterizes our domain.

2.1.2 Mechanism Design for Allocations

The house allocation problem is a fundamental question in mechanism design, tackling the question of how to assign a set of agents one-to-one to a set of heterogeneous items. Hylland and Zeckhauser (1979) study the problem of matching with cardinal preferences and the solution of competitive equilibrium from equal incomes (CEEI). CEEI gives both a natural cardinal notion of efficiency and of fairness. Recently, Alaei et al. (2017) give a polynomial time algorithm for computing the CEEI in matching markets when there are a constant number of distinct agent preferences. To our knowledge, the complexity of computing CEEI in general matching problems is unknown. With linear preferences, but without the unit-demand constraint, CEEI and Nash social welfare coincide and can be computed in polynomial time. Devanur and Kannan (2008) generalize this computational result to piecewise linear concave utilities when the number of goods is constant.

Recently, Budish (2011) considers the generalization from matching to a combinatorial assignment problem where agents may have non-linear preferences over bundles of goods, and shows that an approximate version of CEEI exists. This work also shows that, in large markets, the mechanism that outputs this approximate CEEI is asymptotically truthful. Heuristics for computing the CEEI outcome are given by Othman et al. (2010) and these heuristics have been deployed for the course assignment problem by Budish et al. (2016). On the other hand, Othman et al. (2016) show that the computation of CEEI in these combinatorial assignment problems is PPAD-hard.

The Nash social welfare objective of our work compares to competitive equilibrium from equal incomes of the aforementioned works as follows: the two objectives coincide for linear preferences without the matching constraint (Vazirani, 2007), but with the matching constraint the concepts are not equivalent. Both NSW and CEEI outcomes are Pareto efficient, but to our knowledge, in matching markets, the agents' utilities under the two criteria have not been directly compared. Contrasting with CEEI, for stochastic matchings, the NSW outcomes can be calculated by a convex program, i.e., a program that optimizes the product of utilities over the marginal probabilities given by a doubly-stochastic matrix, and is therefore computationally tractable.

A second line of literature considers ordinal mechanisms for one-sided matching. The *random serial dictatorship* (RSD) mechanism has a long history of practical application. Recently it has been used in applications such as housing and course allocation. Pathak and Sethuraman (2011) study the use of RSD for school choice in New York City. RSD is truthful, ex post Pareto efficient, and easy to implement (e.g., Abdulkadirolu and Sönmez, 1998). On the other hand, RSD is neither ex ante Pareto efficient nor envy-free. To remedy this deficiency of RSD, Bogomolnaia and Moulin (2001) developed the *probabilistic serial* (PS) mechanism which, while not truthful, is ordinally efficient, envy-free, and easy to implement. PS has been studied in various contexts ranging from school assignments to kidney matching and it is often contrasted with RSD. For example, Pathak and Sethuraman (2011) show that students often obtain a more desirable random assignment from PS than from RSD. Nonetheless, under a large market assumption PS and RSD converge and the desirable properties of both are attained (Kojima and Manea, 2010; Che and Kojima, 2010). More recent work has also further studied and compared the performance of these two mechanisms with respect to different metrics both theoretically and experimentally (e.g., Aziz et al., 2016; Hosseini et al., 2018).

Several recent papers have considered approximation in one-sided matching markets without money when agents have cardinal preferences. With cardinal preferences, it is possible to consider the aggregate welfare of an allocation as the sum of the expected utilities of each agent. For an aggregate notion of welfare to make sense, the values of the agents need to be normalized. Two common normalizations are unit-sum, which scales each agent's values so that their sum is one, and unit-range, which scales and shifts each agent's values so that the minimum value is zero and the maximum value is one. Under either of these normalizations, Filos-Ratsikas et al. (2014) show that randomized serial dictatorship is an $\Theta(\sqrt{n})$ approximation and that no algorithm for mapping ordinal preferences to allocations is asymptotically better. Christodoulou et al. (2016) consider the unit-sum normalization and show that the price of anarchy of PS is $\Theta(\sqrt{n})$ and that no mechanism, ordinal or cardinal, is asymptotically better. Important comparison of these above results to ours are as follows: our guarantees do not require a normalization of values. Our approximation guarantees are on per-agent utilities, not on the aggregate welfare which allows some agents to be harmed if other agents benefit. We show that our randomized partial improvement mechanism is asymptotically better than RSD in our per-agent analysis framework by a factor of $\Omega(\sqrt{n})$.

More recently, Immorlica et al. (2017) use a notion of approximate Pareto efficiency to analyze the *raffles mechanism* in one-sided matching markets. This efficiency measure provides per-agent approximation guarantees with respect to the Pareto frontier. Our approximation measure can therefore be thought of as a refinement where instead we compare the agent utilities to a specific highly desired point on the Pareto frontier (the Nash bargaining solution). Instead of eliciting the preferences of the agents, the raffles mechanism instead provides the agents with tickets that they can allocate to items, and items are distributed in proportion to the allocated tickets. As a result, this mechanism is not truthful, but the main result shows that its Nash equilibria are $e/(e-1)$-approximately Pareto efficient, i.e., that there is no equilibrium where each agent's utility is increased by more than an $e/(e-1)$ factor.

The mechanism introduced in Chapter 5 is based on the *partial allocation* (PA) mechanism of Cole et al. (2013) that truthfully and approximately solves the fair division of heterogeneous goods. A novel feature of the PA mechanism is that a fraction of the fair allocation is withheld from individual agents in a way that behaves, in the agents' utilities, as payments that align the incentives of the agents with the Nash social welfare objective. The fair division problem is closely tied to the cake cutting literature, which originated in the social sciences but has garnered interest from computer scientists and mathematicians alike (Abebe et al., 2017; Brams and Taylor, 1996; Moulin, 2003; Robertson and Webb, 1998; Young, 1995). The cake – a heterogeneous, divisible item – is represented by the interval $[0, 1]$ and the agents have valuation functions assigning each subinterval to a non-negative value. These valuations are also assumed to be additive. Algorithmic challenges in cake cutting have recently attracted the attention of computer scientists. A historical overview as well as notable results in cake cutting can be found in surveys by Procaccia (2013) and Procaccia and Moulin (2016). The cardinal matching problem we consider is closely related to the cake cutting problem with piecewise uniform valuations since our agents have linear preferences over items.

Random sampling techniques are now common in the literature on mechanism design. They have been primarily developed for revenue maximization problems where the seller lacks prior information on the agents' preferences (Hartline and Karlin, 2007). Our use of random sampling more closely resembles the literature on redistribution mechanisms, where the designer aims to maximize the consumer surplus and monetary transfers between agents are allowed (Cavallo, 2006; Guo and Conitzer, 2007). An approach by Moulin (2009) is to single out a random agent as the residual claimant, run an efficient mechanism on the remaining agents, and pay the revenue generated by the mechanism to the residual claimant. Similarly, our mechanism randomly partitions the agents into two groups and attempts to implement the PA mechanism on the first group while using the items that would be reserved for the second group to implement the first group's outside option. Further connections between our approach and redistribution mechanisms may be possible.

2.2 Improving Access to Information

2.2.1 Measuring Access to Information

Health information seeking behavior plays a key role in combating the burden of diseases. Online behavior can provide an important lens, especially for stigmatizing conditions (such as STIs), where off-line behavior may be harder to collect (Fox and Duggan, 2013). Health information is central to disease control; for instance, HIV management requires extensive informational support to maintain the well-being of those affected and their caretakers. However, there is inadequate understanding of individuals' health information seeking behavior, disease knowledge, and perceptions for the three diseases of interest in this paper (Chan and Tsai, 2018; Hogan and Palmer, 2005).

This lack of knowledge is especially prominent for individuals living in developing nations. There are relatively few studies, and existing studies are often limited to specific subpopulations. For instance, Abimanyi-Ochom et al. (2017) explore the HIV/AIDS knowledge, attitudes, and practices of Ugandan individuals with disabilities. Similarly, Gombachika et al. (2013) set out to understand how couples living with HIV in Malawi obtain sources of information and reproductive decisions. Studies covering all 54 nations in the continent have often focused on aggregate health outcomes, such as quantifying the burden of diseases by country, rather than health information consumption.

Search and other large-scale Web data have emerged as key for understanding health patterns, health information consumption, and characterizing communities facing certain health outcomes (Abebe et al., a; De Choudhury and De, 2014; Fox and Duggan, 2013; Sillence et al., 2007; Liu et al., 2013; Eysenbach and Köhler, 2002; Spink et al., 2004; De Choudhury et al., 2014). Online health information seeking behavior is known to be connected to off-line behavior and can inform health policy (Fiksdal et al., 2014; Ling and Lee, 2016; Zheluk et al., 2013; Ocampo et al., 2013; O'Grady, 2008). Search data is especially valuable as it is real-time, detailed, relevant, and gives less-filtered insights into individuals' health information needs (Eysenbach and Köhler, 2004).

Despite these findings, studies focusing on the use of search engines as a medium for obtaining health information related to the three diseases have remained small scale, often limited to surveys and focused on developed nations (O'Grady, 2008; Hogan and Palmer, 2005; Shuyler and Knight, 2003). There is need to understand individuals' search behavior before attempting to target relevant information to individuals whether on- or off-line (Fiksdal et al., 2014).

A closely related line of work to ours is surveillance and case finding, where there is extensive work related to HIV/AIDS, malaria, and tuberculosis (Zhou and Shen, 2010; Ocampo et al., 2013). This work shows promising results connecting on- and off-line behavior and suggests that search data can be valuable as an information source for health. Similar work related to other diseases include using search data for influenza outbreaks (Ginsberg et al., 2009; Polgreen et al., 2008; Santillana et al., 2015; Yang et al., 2015; Yuan et al., 2013), dengue fever (Althouse et al., 2011; Chan et al., 2011), norovirus outbreaks (Desai et al., 2012), bacterial infections (CDC, 1998), and many other outbreaks (Brownstein et al., 2009; Carneiro and Mylonakis, 2009; Hay et al., 2013; Rothman et al., 2008).

Health outcomes can vary drastically by demographics, especially in developing nations. Gender and age impact likelihood of being infected and ability to obtain care and treatment. For instance, it is known that 61% of all sub-Saharan individuals living with HIV are women, and women in the 15–24 age group are three times more likely than men in the same age group to acquire HIV (WHO, 2009). Men are more likely to develop and die from tuberculosis (UNDP, 2015b). Pregnant women are disproportionately impacted by malaria (UNDP, 2015a).

This variance extends to individuals' knowledge about the diseases of interest (Fransen-dos Santos, 2009). Young individuals are evaluated to have incomprehensive knowledge about these diseases, and young women especially so (Stonbraker et al., 2017; WHO, 2009; Li et al., 2004; Kumar and Mmari, 2004; Wang et al., 2008). At the same time, it is difficult to find large age and gender-disaggregated data as they are not routinely collected or reported (UNDP, 2015a,b). This highlights a research gap and potential for computationally-informed policy contributions for effective prevention, coverage, and treatment of these diseases. In resource-constrained settings where funds must be allocated strategically, it is especially prudent to take the varied needs of these groups into account.

2.2.2 Network-Based Interventions

Word-of-Mouth Recommendations

A popular phenomenon in social networks is word-of-mouth recommendations, in which information spreads from one person to another, potentially spreading across the full network (Domingos and Richardson, 2001; Kempe et al., 2003; Richardson and Domingos, 2002). *Influence maximization*, the process of selecting the most influential agents on a network in order to propagate information across the full population, is one popular strategy which has been employed in a variety of contexts including in health (Wakefield et al., 2003; Yadav et al., 2016, 2017, 2018). There has been some theoretical work showing the counter-intuitive outcome where increased effort in influence maximization results in a less successful spread. An example is Sela et al. (2016), where they show that due to the separation of the infection and viral stage, there are cases where an increased effort can result in a lower rate of spread. A related line of work has made use of rich data sets on digital friend-to-friend recommendations on e-commerce sites to analyze the flow of innovation recommendations through an underlying social network (Leskovec et al., 2007a). Further work has experimentally explored influence strategies, with individuals either immediately broadcasting their innovation adoption to their social network, or selecting individuals to recommend the innovation to (Aral and Walker, 2011).

The consequences of negative reactions from agents on a network have been analyzed in a range of different domains. In the introduction we noted examples involving Groupon (Byers et al., 2012), book prizes (Kovcs and Sharkey, 2014), and on-line media collections (Aizen et al., 2004). Although experimentally introduced negative ratings tend to be compensated for in later reviews, positive reviews can lead to herding effects (Muchnik et al., 2013). There has also been research seeking to quantify the economic impact of negative ratings, in contexts ranging from seller reputations in on-line auctions (Bajari and Hortascu, 2004; Resnick et al., 2006) to on-line product reviews (Pang and Lee, 2008). This work has been consistent in ascribing non-trivial economic consequences to negative impressions and their articulation through on-line ratings and reviews. Recent work has also considered the rate at which social-media content receives "likes" as a fraction of its total impressions, for quantifying a social media audience's response to cascading content (Rotabi et al., 2017).

The literature on pricing goods with network effects is another domain that has developed models in which consumers are heterogeneous in their response to diffusing content. The underlying models are different from what we pursue in Chapter 8; a canonical structure in the literature on pricing with network effects is a set of consumers with different levels of willingness to pay for a product (Katz and Shapiro, 1985). This willingness to pay can change as the product becomes more popular; a line of work has thus considered how a product with network effects can be priced adaptively over time as it diffuses through the network (Arthur et al., 2009; Hartline et al., 2008). The variation in willingness to pay can be viewed as a type of "criticality," with some consumers evaluating products more strictly and others less strictly. But a key contrast with our work is that highly critical individuals in these pricing models do not generally confer a negative payoff when they refuse to purchase an item.

Network Segregation and Inequality

Homophily is a robust and prevalent process impacting network formation in many domains (Lazarsfeld et al., 1954; McPherson et al., 2001; Newman, 2002). There is a long line of theoretical and empirical work exploring the effect of homophily on network formation, ranging from observational studies on large network data, to laboratory experiments, to analyses of theoretical models (Adamic and Glance, 2005; Dong et al., 2017; Goeree et al., 2009). A main topic of focus has been the interaction between homophily and network segregation. In theoretical work, Currarini et al. (2009); Henry et al. (2011) show that segregated networks emerge due to homophily. In related work, Kim and Altmann (2017) study the effect of homophily on the rich-get-richer phenomena. Empirical work such as by Mayer and Puller (2008) has explored the effect of homophily on integration in settings like college campuses. In related work to ours, Bramoullé et al. (2012) adapt the Jackson-Rogers model to the case with heterogeneous nodes. Their work primarily focuses on how each node's likelihood to form links changes over time; in contrast, we consider a global measurement of integration, using the fraction of bichromatic edges.

Triadic closure is another well-studied process in network formation dating back to the 1950s (Kossinets and Watts, 2006; Rapoport, 1953). While there is a long line of work on the effect of triadic closure on network clustering, the interplay between homophily and triadic closure remains under-explored. In one tangentially related work, Altenburger and Ugander (2018) show that monophily – the presence of individuals with preference for attributes unrelated to their own – has a tendency to induce similarity among friends-of-friends. In contrast, in our work, we study the relationship of homophily and triadic closure. Nevertheless, some of the insights here complement ours.

Segregation in social networks can limit individuals' ability to access to opportunities, leading to the creation or exacerbation of disparities across groups. Research across various disciplines has modeled and measured the impact of segregation on social welfare including its impacts on access to information, economic development, educational outcomes, labor market outcomes, and social capital and support, (Banerjee et al., 2013; Calvo-Armengol and Jackson, 2004; Calvo-Armengol et al., 2009; Dasaratha, 2017; DiMaggio and Garip, 2011; Eagle et al., 2010; Jackson et al., 2012; Banerjee et al., 2013). Recent work, such as by Avin et al. (2015) has proposed and studied models that explain how inequality and disparities in access to opportunity arise in certain settings.

Our work has additional implications for network-based interventions both in on- and off-line settings. For instance, it is well-known that biases that may exist on online platforms such as Twitter and Task Rabbit may lead to inequalities between groups leading inequalities between groups (Nilizadeh et al., 2016; Hannák et al., 2017). These biases , amplified by recommendation algorithms, can impact how networks grow and evolve creating an *algorithmic glass ceiling* (Biega et al., 2018; Stoica et al., 2018; Su et al., 2016; Biega et al., 2018).

In recent years, there has been interest by researchers in fairness in recommender systems concerning the effect of small nudges on the long-term health of the platform such as by mitigating segregation, improving interactions, and achieving other desirable societal objectives (Ekstrand and Willemsen, 2016; Guy, 2015; Hutson et al., 2018; Knijnenburg et al., 2016; Schnabel et al., 2018; Stoica et al., 2018; Su et al., 2016). These studies have shown that the platform designer, by using small nudges when the user first joins, may be able to realize large gains over time. The intervention considered in Chapter 9 shows that in settings where there is an underlying network structure governing interactions, small interventions when a user first joins the platform may be further amplified by triadic closure, resulting in a more integrated network over time. Such interventions mirror ones similar to dorm assignments with an eye to integration on college campuses.

Part II

Societal Resource Allocations

CHAPTER 3

OVERVIEW OF PART II

In Part II, we discuss what role algorithms and mechanism design can play in informing allocation of societal resources. We are specifically inspired by settings where such resources may be scarce or where there is heterogeneity in families' or individuals' need or desire for these resources.

As discussed in Chapter 1, the works presented here highlight two sets of challenges: in the first section, we investigate the problem of modeling families' welfare informed by empirical work. Specifically, in Chapter 4, we focus on the question of modeling welfare when families are experiencing *income shocks*. We use this model to pose a set of optimization-based questions for mitigating undesirable outcomes such as likelihood of eviction. We present optimal and near-optimal algorithms for allocating subsidies in a variety of general settings. We further show that the structure of these solutions give further insight into the nature of economic welfare and discuss their societal implications.

The allocation of resources – such as subsidies, low-income housing resources, or seats in public schools – often pose technical challenges and give rise to fundamental mathematical questions. One such question is the *house allocation* problem where a set of individuals must be matched one-to-one to a set of items (such as houses). In Chapter 5, we present a mechanism for this allocation problem satisfying several desirable fairness and efficiency criteria that were not previously simultaneously satisfied by a single mechanism. We provide theoretical and experimental evidence that this mechanism outperforms several popular ordinal mechanisms used in this setting.

CHAPTER 4

SUBSIDY ALLOCATIONS IN THE PRESENCE OF INCOME SHOCKS

Understanding and measuring economic hardship is a fundamental question that directly informs the design of policies and assistance programs aimed at addressing the needs of vulnerable individuals and families (Alkire and Foster, 2011; Anand and Sen, 1997; Atkinson, 2003; Council et al., 2011; Yapa, 1996). A crucial challenge in this activity is the range of factors that play a role in poverty and economic hardship, including health, demographics, social ties, intergenerational dynamics, and many other dimensions (Grusky, 2018; Grusky and Ku, 2008; Grusky and Weeden, 2016).

Recent studies have sought to address the gap between official measures of welfare and the more complex formulations that might be needed to accurately identify the sources of greatest need. One active and ongoing effort is the *Poverty Tracker* program (Lens et al., 2016; Wimer et al., 2014a, 2016a, 2014b, 2016b), which surveys approximately 2300 families in New York City and documents the intricate associations between their circumstances and levels of hardship.[1] As with other studies in this area, it is in part based on the premise that we need to broaden our frameworks for quantifying economic well-being; the researchers involved in the program write that,

> ...official 'income only' measurements of poverty ... painted a picture that was too optimistic and didn't capture the magnitude of disadvantage, nor the true struggles New Yorkers face in trying to make ends meet (Wimer et al., 2014a).

[1] The Poverty Tracker study is based on surveys of families of all income levels across New York City and is conducted by the Robin Hood Foundation and the Columbia University Population Research Center.

So, what are the missing dimensions in these basic measurements?

Several additional components of economic hardship manifest themselves through a common mechanism: unexpected and disruptive *shocks* to a family's economic state (Giesbert and Schindler, 2012; Kijima et al., 2006; Kochar, 1995; Morduch, 1994). Such shocks can be a result of an unexpected expense (e.g., a parking ticket or a health bill caused by an accident), a delayed paycheck or unexpected loss of a job, a dissolution or loss of a romantic or other close personal relationship, interactions with the criminal justice system, and many other experiences (Lens et al., 2016; Mello, 2018; Wimer et al., 2014a, 2016a, 2014b, 2016b). Income shocks are receiving increasing attention from social scientists and policy-makers alike; summarizing a recent round of findings, the Poverty Tracker analysis discussed above reports that "the most persistently disadvantaged New Yorkers are beset by repeated shocks to their finances and well-being" (Wimer et al., 2016a).

A crucial point is that families vary significantly in their ability to withstand income shocks; while an unexpected bill might be a mere inconvenience for some families, for other families, it can lead to eviction, poor health, loss of a job, and other undesirable outcomes that may trigger or lock families into persistent poverty (Atake, 2018; Björkman-Nyqvist, 2013; Desmond, 2012, 2015; Desmond and Shollenberger, 2015; Desmond, 2016; Desmond and Gershenson, 2016; Dinkelman et al., 2008; Kochar, 1999; O'Flaherty, 2009). In many cases, it is significantly more challenging to remedy the consequences of such experiences than it is to prevent families from experiencing them in the first place. For instance, a common consequence of inability to withstand income shocks is eviction, which has been argued to be a leading cause of poverty (Desmond, 2012, 2016). The *Milwaukee Area Renters Study* (MARS) shows that a recurring reason why families fail to pay their bills month-to-month is inability to recover from unexpected shocks in the form of an unexpected bill or a disruption to their income flow.[2] The disruption to families' lives caused by eviction has been extensively documented to negatively impact families' well-being.

We thus return to the disconnect noted at the outset: despite the centrality of shocks in hardship, they do not play a correspondingly central role in the evaluations and decisions made by social assistance programs. For instance, standard eligibility guidelines for housing assistance programs are based on income, adjusted for family size, or percentage of income spent on housing. Likewise, other assistance programs such as the Supplemental Nutrition Assistance Program (SNAP), Medicaid, and Low Income Home Energy Assistance Program (LIHEAP) are based on income eligibility. Yet two families that look similar under such measures may still differ significantly in their vulnerability if one family experiences a significantly different profile of shocks.

There is thus a danger of misprioritization, if we are not taking into account factors that we know to be crucial. What would it look like to incorporate information about shocks into disadvantage determinations, and how might it inform decisions about assistance? There is a long line of work in the study of household consumptions and public economics that considers agent-level behavior in response to subsidies and more broadly understanding consumption dynamics (Aiyagari, 1994; Golosov et al., 2006; Kocherlakota, 2004) as well as optimal taxation theory (Eaton and Rosen, 1980a,b,c; Farhi and Werning, 2013; Saez and Stantcheva, 2018; Varian, 1980). The work presented here is informed

[2]The Milwaukee Area Renters Study is based on surveys administered by the University of Wisconsin Survey Center of over 1000 low-income households who rented in the Milwaukee private housing market in 2009-2011.

by these areas. We however take an algorithmic approach and study the inherent stochastic optimization problems at the heart of subsidy allocation with shocks.

Our Results

In this chapter, we develop a stylized model for the state of an agent (representing a family potentially in need of assistance) as they experience shocks over time. We then use this model to formulate the problem of allocating subsidies to agents, when the total amount of subsidy is constrained by a given budget. One option for approaching this problem is to take into account only the income of each agent. But, given that shocks can have a significant effect on families' welfare, how should we incorporate information about shocks into the search for allocations? And, how much can our allocation decisions change when we use this information about shocks?

We focus on the algorithmic issues inherent in the stochastic optimization problem at the heart of subsidy allocation with shocks. Our model is based on the framework of *ruin probabilities* from the literature on stochastic risk modeling (Asmussen and Albrecher, 2010). We assume that each agent's state is represented by a numerical *reserve*, corresponding to the current buffer they have for absorbing shocks. An agent receives a flow of income continuously over time, which increases their reserve, and at discrete points in time they can experience a shock, which abruptly decreases their reserve by a given amount. When the reserve becomes negative (or more generally, drops below some designated threshold), the agent experiences a significant undesirable life event, such as eviction. We will refer to this as a state of *ruin*; since remedying the consequences of such ruin events can be profoundly greater than preventing the occurrence in the first place, we would like to keep such events from happening.

To help prevent agents from experiencing ruin, we assume that we have a fixed set of funds that we can allocate across the set of agents. We consider models in which these funds can be used to supplement an agent's rate of income flow (via an *income subsidy*), or to directly supplement their initial wealth (via a *wealth subsidy*), or both. For a set of different objective functions — for example, to minimize the expected number of agents that experience ruin (min-sum objective) or to minimize the maximum ruin probability experienced by any agent (min-max objective) — we study the problem of optimally allocating subsidies to achieve this objective.

We obtain three sets of results for this problem:

- We obtain optimal allocation algorithms for the min-max objective in the general setting.
- We present optimal allocation algorithms for the min-sum objective for two natural special cases: first, when the initial wealth of each agent is zero, and second, when the sizes of shocks are exponentially distributed.
- For the min-sum objective in the general setting, we give a fully polynomial-time approximation scheme to the optimal solution.

Note that, for the last result, our approximation scheme works even for shock distributions that have no compact representation, and are available to us only in a "black-box" fashion.

In the process, we obtain a number of insights into the structure of the problem. In particular, the efficient algorithms we obtain turn out to have a natural structure based on *priority orderings* on the agents: roughly speaking, they order the agents by a numerical measure of responsiveness to subsidy, and then begin allocating subsidies in decreasing order of responsiveness. We believe that these priority orderings are interesting objects in and of themselves: they provide a natural way of quantifying how each agent would respond to assistance; and since the orderings do not simply follow the sorted order of income or wealth reserve, they highlight the ways in which the presence of shocks ought to change our reasoning about the allocation of assistance. Moreover, the optimal allocations also look quite different depending on whether we are providing income subsidies or wealth subsidies, suggesting that the nature of the intervention can have a significant effect on how we prioritize agents in need of help.

4.1 Allocating Wealth and Income Subsidies

We begin by specifying the theoretical model and problem formulation. The model is based on the structure of *ruin processes* that are standardly used to represent risk in insurance markets (Asmussen and Albrecher, 2010); to make the exposition self-contained, it is useful to describe our version of the model from first principles. The optimization problems we study, based on interventions to modify the ruin probabilities, are natural given the motivation in the previous section, but less standard in the earlier literature as it is focused on insurance markets rather than poverty-reduction. We describe these problems in the latter part of the section, after first establishing the basic model.

A Model of Income, Reserve, and Shocks. There are n agents; we can think of each as representing a family that a planner would like to assist. Each agent i has a net income c_i per unit time, and an initial reserve u_i. What we call (net) income here can be thought of as the difference between the agent's earned income minus their expenses during each time period. Time runs continuously; so, in the absence of shocks, agent i's reserve after an amount of time t would be $u_i + c_i t$.

The shocks experienced by agent i operate as follows: shocks arrive at randomly selected discrete points in time $T_{i1}, T_{i2}, \ldots$ with the gaps between them $T_{i(j+1)} - T_{ij}$ distributed independently according to some distribution G_i. Thus, if a shock happens at some time T, we can imagine setting an independent random "countdown timer" of length drawn from G_i; when this timer expires, the next shock happens. When the j^{th} shock happens, it has a size S_{ij} drawn from a *shock-size distribution* F_i, and S_{ij} is subtracted from the agent's current reserve.

For concreteness, unless stated otherwise, assume that the shocks arrive according to a *Poisson process*, which has the structure described above, with the gaps between consecutive shocks drawn from a distribution G_i that is exponentially distributed with rate β_i. The expected length of the gap between consecutive shocks is $1/\beta_i$; we can equivalently think of this as saying that there are β_i shocks per unit time in expectation. The use of an exponential distribution for the gap between consecutive shocks yields the so-called *Cramer-Lundberg model* from the theory of ruin processes (Asmussen and Albrecher, 2010). We note that the use of the exponential distribution will be important for the special cases we study; but our results on general distributions extend to an essentially arbitrary gap distribution G_i.

In summary, the agent's reserve at time t, denoted $R_i(t)$, is given by the equation

$$R_i(t) = u_i + c_i t - \sum_{j:T_{ij} \leq t} S_{ij},$$

where the last term is simply the total size of all shocks that have arrived by time t. (The number of shocks arriving by any finite time t is finite in the model with probability 1.)

Ruin Probability. Our goal is to help agents keep their reserve from becoming negative; if $R_i(t) < 0$ at any time t, then we say the agent has experienced *ruin*. Let ψ_i be the probability that there exists a time t at which agent i experiences ruin; since this is a function determined entirely by the agent's income c_i, initial wealth u_i, arrival rate of shocks β_i, and shock-size distribution F_i, we can write it as $\psi_i = \psi(c_i, u_i, \beta_i, F_i)$.

The qualitative behavior of the ruin probability ψ_i depends heavily on a parameter called *drift*, which captures the expected change per unit time in the agent's reserve. Specifically, if the expected value of the shock-size distribution is μ_i, then the drift is equal to $c_i - \beta_i \mu_i$. A standard result is that if the drift is negative — so that the agent's reserve is being pulled downwards in expectation — then the ruin probability ψ_i is equal to 1: the agent will be ruined almost surely as t goes to infinity. On the other hand, if the drift is positive, there is still a chance that agent can be ruined by shocks that are large and/or rapid enough; but it can be shown that $\psi_i < 1$, so there is a positive probability that the agent will never be ruined even as time runs to infinity (Asmussen and Albrecher, 2010). In the special cases we study, we focus on the case of positive drift, where the agents might be able to avoid ruin on their own, but we would like to help lower their ruin probabilities. When we move to the general case, we will allow both positive and negative drift.

Optimization. We now consider how to model the problem of providing assistance to the agents. Let us first consider the case of *income subsidies*, in which we have a budget B of funds, and we can choose to increase the income of agent i by an amount x_i, as long as the total amount $\sum_{i=1}^{n} x_i$ that is given out is at most B. We would like to do this so as to reduce some objective function based on the ruin probabilities of the agents. The choice of objective function reflects a societal preference on which outcomes are most desirable; to be concrete, we observe that two natural objectives are a *min-sum* formulation and a *min-max* formulation.

In the min-sum formulation, each agent i has a weight w_i representing the social cost resulting from ruin to agent i. The differences in these weights — corresponding to differences in social cost — may reflect the fact the loss of a job, eviction, or other economic hardship can impact different families differently (Desmond, 2012, 2016; Shapiro et al., 2004). These differences could come from the dynamics or composition of the family; or the fact that some families might have outside options – such as economically stable extended families or the ability to take out loans – while for other families the ruin event might trigger a cycle of poverty from which they cannot escape.

With these weights in place, the min-sum objective seeks to minimize the weighted expected number of agents who experience ruin. In the notation we have developed so far, the goal is then

$$\min_{x_1+\cdots+x_n=B} \sum_{i=1}^{n} w_i \psi(c_i + x_i, u_i, \beta_i, F_i),$$

where we observe that $\psi(c_i + x_i, u_i, \beta_i, F_i)$ denotes the ruin probability of agent i

after a subsidy of x_i has been added to their income.

In contrast, the min-max formulation seeks to ensure that the worst ruin probability experienced by any agent is as low as possible. Thus, the goal is

$$\min_{x_1+\cdots+x_n=B} \max_{i=1,\ldots,n} \psi(c_i + x_i, u_i, \beta_i, F_i).$$

These functions correspond to two well-studied societal objectives. Of course, these are not the only two reasonable objective functions. Societal implications in this choice are discussed in Section 4.5 and Part IV, where we more generally discuss where and how this process of formalizing problems can serve as an ally for desirable societal change.

Instead of an income subsidy, we could alternatively consider a *wealth subsidy*: an amount z_i is added to agent i's initial wealth so as to reduce the ruin probability. We can again formulate min-sum and min-max versions of the problem with wealth subsidies; the only difference at the level of the notation set-up is that the ruin probability of agent i with wealth subsidy z_i is evaluated as $\psi(c_i, u_i + z_i, \beta_i, F_i)$. We can also study a mixed income and wealth subsidy, in which there is a "conversion factor" k between money allocated for wealth subsidies and income subsidies; the planner solving the optimization problem can give income subsidy x_i and wealth subsidy z_i to agent i, provided that $\sum_i z_i + k \sum_i x_i \leq B$. Note that although a wealth subsidy is implemented as a one-time allotment while an income subsidy is paid out over time, it is natural to convert between them if we take the standard interpretation of income subsidies as coming from a fixed endowment that grows geometrically to support a fixed rate of payout.

Overview of Main Questions. As discussed, assistance programs generally do not take shocks into account, and instead use some function of income or reserve. Our model allows us a natural way to incorporate shocks into decisions about allocating assistance. We therefore set out to answer the following questions: first, are there efficient algorithms to optimize the objective functions based on the min-sum and min-max formulations of assistance described in this section? For this question, as noted in the introduction, we give efficient algorithms in the general setting for the min-max objective and for some natural special cases for the min-sum objective. The algorithms in these cases are based on priority orderings of the agents by need, based on explicit functions of their income, initial wealth, shock rate, and shock-size distribution. For the general case, we give a fully polynomial-time approximation scheme for the min-sum objective.

Given the ordering structure inherent in these solutions, we also ask how different these orderings are from simple sortings of the agents by income, initial wealth, or even ruin probability. We show that the optimal priority orderings can be different from all of these; and in fact, in some examples, they can be the exact reverse of each these simpler orderings. Additionally, the optimal priority ordering for income subsidies can, in some cases, be the exact reverse of the optimal priority ordering for wealth subsidies. That is, the strategy one should use when income subsidies are the planned intervention may be very different from the strategy one should use with wealth subsidies. We now turn our attention to a special case of our problem.

4.2 Agents with Zero Initial Wealth

We would like to characterize subsidies that either minimize the weighted expected number of agents who experience ruin (the *min-sum objective*) or the largest ruin probability of any agent (the *min-max objective*). We consider this first for the fundamental case of agents who have no initial wealth. This is a natural instance of the problem to explore given that the empirical and policy work in these domains generally focuses on instances where individuals have almost no existing buffer against ruin. Assuming an initial wealth equal to 0 is an abstraction of this challenging case.

As before, each agent i is characterized by an income c_i, an initial wealth u_i, which is equal to 0 in the present case, and shocks that arrive according to a Poisson process of rate β_i, and with sizes drawn from a distribution F_i. We are interested in the probability that agent i will experience eventual ruin; this is given by $\psi_i = \psi(c_i, 0, \beta_i, F_i)$.

As is standard in the theory of ruin processes, we will make the following mild assumption about the shock-size distribution F_i throughout the section — that if we let Z_t denote the random variable equal to the total size of all shocks occurring between times 0 and t, then the quantity Z_t/t (the average amount of shock per unit time) converges to a constant limit with probability 1 (Asmussen and Albrecher, 2010). This condition is satisfied whenever the shock-size/shock-arrival distributions have finite mean and variance, and therefore essentially all distributions we might wish to consider.

A fundamental result in the theory of ruin processes is that when the initial wealth is 0, the ruin probability ψ_i depends on the shock-size distribution F_i

only through its mean value μ_i: if F_i and F'_i are shock-size distributions with the same means, then $\psi(c_i, 0, \beta_i, F_i) = \psi(c_i, 0, \beta_i, F'_i)$ (Asmussen and Albrecher, 2010). Thus, in a mild extension of our notation, we will write $\psi(c_i, 0, \beta_i, \mu_i)$ to stand in for $\psi(c_i, 0, \beta_i, F_i)$, when μ_i is the mean of F_i. Moreover, the ruin probability has a particular clean functional form given by

$$\psi(c_i, 0, \beta_i, \mu_i) = \frac{\beta_i \mu_i}{c_i}. \tag{4.1}$$

Here, we will focus on income subsidies: we would like to assign an income subsidy x_i to each agent i, thus reducing their ruin probability from $\frac{\beta_i \mu_i}{c_i}$ to $\frac{\beta_i \mu_i}{c_i + x_i}$, in such a way that the sum of the x_i is constrained by the overall budget B.

To analyze this process in terms of the min-max and min-sum objectives, it is useful to formulate the underlying optimization problem more abstractly. This abstraction will be useful for other special cases we consider as well as the generalized version of our problem. We do this next, before returning to the application for agents with zero initial wealth.

4.2.1 An Abstract Formulation

There is an abstract optimization problem that will provide a useful unifying description for the current problem and several of the subsequent ones we consider. The problem and its solution are related to "water-filling algorithms" from the theory of convex minimization (Alaei et al., 2012; Boyd and Vandenberghe, 2004); we describe it here because the form of the solution provides an important structural insight for our domain — that the optimal allocation of subsidies in each case is based on a *priority ordering* of the agents. We first describe this abstract problem and its solution, and then we show how it applies to agents with zero initial wealth. Our problem is as follows:

> ($*$) *We have functions $f_1, \ldots, f_n$. Each f_i is a continuous function of a single real-variable that is positive and strictly decreasing: if $x < z$ then $f_i(x) > f_i(z)$. We would like to find non-negative real numbers $x_1, \ldots, x_n$ so that $\sum_{i=1}^{n} x_i = B$ and $\max_i f_i(x_i)$ is minimized.*

Intuitively, if we think of f_i as the ruin probability of agent i when given income subsidy x_i, we see that the min-max objective is a direct special case of Problem ($*$). But, as we will see later in the section, this formulation will allow us to solve the min-sum objective as well.

It is useful to first discuss the special case of Problem ($*$) when all $f_i(0)$ are the same.

Lemma 1. *If $f_i(0) = f_j(0)$ for all i and j, then there is a unique vector $x^* = (x_1^*, \ldots, x_n^*)$ with the property that $\sum_i x_i^* = B$ and $f_i(x_i^*) = f_j(x_j^*)$ for all i and j. Moreover, x^* uniquely optimizes Problem ($*$) in this case.*

Proof. First, since $\max_i f_i(x_i)$ is a continuous function on the compact set defined by $\sum_i x_i = B$ and $x_i \geq 0$ for all i, it achieves its minimum at some vector $x^* = (x_1^*, \ldots, x_n^*)$.

Let q be the common value of all $f_i(0)$. We observe that the optimal value of $\max_i f_i(x_i)$ must be strictly less than q, since setting $x_i = B/n$ satisfies $\max_i f_i(x_i) < q$ by the strict decreasing property of all f_i. Thus, we must have $x_i^* > 0$ for all i, since otherwise would have $\max_i f_i(x_i^*) = q$, contradicting the optimality of x^*.

We claim that x^* satisfies the desired condition $f_i(x_i^*) = f_j(x_j^*)$ for all i and j. For, if not, let S be the set of indices i for which $f_i(x_i^*)$ is maximal. Since we do not have $f_i(x_i^*) = f_j(x_j^*)$ for all i and j, there is some index h that does not belong to S. But now, by the continuity and strict decreasing property of the functions f_i, we can choose a sufficiently small $\varepsilon > 0$, increase x_i^* by ε for each $i \in S$, and (since $x_h^* > 0$) decrease x_h^* by $\varepsilon|S|$, so as to reduce the value of $\max_i f_i(x_i)$. This contradiction shows that we must have $f_i(x_i^*) = f_j(x_j^*)$ for all i and j.

To show uniqueness, suppose there were two distinct vectors x^* and z^* that both satisfied the equality $f_i(x_i^*) = f_j(x_j^*)$ and $f_i(z_i^*) = f_j(z_j^*)$ for all i, j. Then, for some i, we must have $x_i^* \neq z_i^*$; suppose (by symmetry) that $x_i^* > z_i^*$. Since f_i is strictly decreasing, it follows that $f_i(x_i^*) < f_i(z_i^*)$. Now, since all f_j are strictly decreasing, we would then have $f_j(x_j^*) < f_j(z_j^*)$ for all j, from which it follows that $x_j^* > z_j^*$ for all j. But, this contradicts the assumption that both $\sum_i x_i^*$ and $\sum_i z_i^*$ are equal to B. $\square$

Building on this special case, we would like to study the behavior of the following *priority algorithm* for solving Problem $(*)$. First, we index the functions so that $f_1(0) \geq f_2(0) \geq \cdots \geq f_n(0)$; that is, if $i \leq j$, then $f_i(0) \geq f_j(0)$. The algorithm is then easy to describe informally: we increase x_1 continuously until $f_1(x_1)$ matches $f_2(0)$; we then continuously increase both x_1 and x_2 simultaneously, keeping the values of $f_1(x_1)$ and $f_2(x_2)$ equal to each other, until they both match $f_3(0)$; we then continuously increase x_1, x_2, x_3 simultaneously, keeping their values equal to each other, until they all match $f_4(0)$; and, we proceed in this way until we reach the budget B.

We can describe the algorithm more formally as follows:

- For each $i \le j$, we find the value δ_{ij} such that $f_i(\delta_{ij}) = f_j(0)$: that is, the input we can provide to f_i if we want to reduce its value to $f_j(0)$. Since each function is strictly decreasing, there is a unique such δ_{ij}.
- We now find the maximum j such that $\sum_{i<j} \delta_{ij} \le B$. Let m be this value of j; note that $m \ge 1$. (For notational simplicity, we define $\delta_{mm} = 0$.)
- We are now at a value of m such that if we tried to increase each of $x_1, \ldots, x_m$ so that $f_i(x_i) = f_{m+1}(0)$, we would exceed our budget B.
- If $\sum_{i<m} \delta_{im} = B$, then we declare this to be our solution: we set $x_i = \delta_{im}$ for $i < m$, and $x_i = 0$ for $i \ge m$.
- Otherwise, if $\sum_{i<m} \delta_{im} = C < B$, then we apply Lemma 1 as follows. We define a function $g_i(z_i) = f_i(\delta_{im} + z_i)$ for $i \le m$, and then we apply Lemma 1 to the functions $g_1, \ldots, g_m$ with budget $B - C$. This produces a unique vector $z^* = (z_1^*, \ldots, z_m^*)$. As our solution, we report $x_i = \delta_{im} + z_i^*$ for $i \le m$, and $x_i = 0$ for $i > m$.

We think of this as a *priority algorithm* because it arrives at a solution by increasing the values of x_i in a natural priority ordering: first just x_1, then both x_1 and x_2 simultaneously, and so forth.

We now establish the basic properties of the solution returned by this priority algorithm.

Theorem 2. *Let $x^* = (x_1^*, \ldots, x_n^*)$ be the solution returned by the priority algorithm. Then x^* is the unique vector that minimizes the objective function $\max_i f_i(x_i)$, and it is also the unique vector satisfying the following property:*

> *(†) (i) If x_i and x_j are both positive, then $f_i(x_i) = f_j(x_j)$; and (ii) if $x_i > 0$ but $x_j = 0$, then $f_i(x_i) \ge f_j(x_j)$.*

Proof. Let $f^* = \max_i f_i(x_i^*)$. If $v = (v_1, \ldots, v_n)$ is any vector whose coordinates sum to B but is not equal to x^*, then we have $v_i < x_i^*$ for some index i. Let m be the index computed by the priority algorithm. The vector x^* only has positive coordinates for indices $h \leq m$, and $f_h(x_h^*) = f^*$ for all such indices. Since f_i is strictly decreasing, we thus have $f_i(v_i) > f_i(x_i^*) = f^*$, and hence v cannot be optimal. Thus, x^* is the unique vector that minimizes the objective function.

It must also hold that $v_j > x_j^*$ for some index j. Then, $f_j(v_j) < f^*$. Since x^* is optimal, there must also be an index i for which $f_i(v_i) \geq f^*$. Either $v_i > 0$, which violates part (i) of (†), or $v_i = 0$, which violates part (ii) of (†). Thus, x^* is the unique vector that satisfies (†). □

4.2.2 The Case of Zero Initial Wealth

We now return to our motivating question — how to optimally allocate income subsidies to agents with zero initial wealth. As before, the ruin probability of agent i, with income c_i and shocks of arrival rate β_i and mean size μ_i, is given by $\psi(c_i, 0, \beta_i, \mu_i) = \frac{\beta_i \mu_i}{c_i}$.

For the min-max objective, we can directly apply the priority algorithm developed for the formulation in the preceding subsection. We define $f_i(x_i) = \psi(c_i + x_i, 0, \beta_i, \mu_i) = \frac{\beta_i \mu_i}{c_i + x_i}$, and we find a subsidy to minimize $\max_i f_i(x_i)$. The priority ordering used by the algorithm in this case is the ruin probability itself, $f_i(0) = \psi(c_i, 0, \beta_i, \mu_i)$.

For the min-sum objective, we will also be able to use the abstract formulation as follows. Let the objective function in the min-sum case by denoted by a function ϕ where

$$\phi(x_1, \ldots, x_n) = \sum_{i=1}^{n} w_i \psi(c_i + x_i, 0, \beta_i, \mu_i).$$

Using Equation 4.1, we can restate this to be

$$\phi(x_1, \ldots, x_n) = \sum_i w_i \frac{\beta_i \mu_i}{c_i + x_i}.$$

We now take the partial derivative of ϕ with respect to x_i to get

$$\frac{\partial \phi}{\partial x_i} = -\frac{w_i \beta_i \mu_i}{(c_i + x_i)^2}. \tag{4.2}$$

We see that ϕ is strictly convex by taking the second partial derivative with respect to x_i

$$\frac{\partial^2 \phi}{\partial^2 x_i} = \frac{2 w_i \beta_i \mu_i}{(c_i + x_i)^3},$$

which is strictly positive for all $x_i \geq 0$.

Because of the strict convexity, ϕ has a unique local (and hence also global) minimum over the set defined by $x_i \geq 0$ and $\sum_i x_i = B$. We can characterize it using the priority algorithm from our abstract formulation as follows: we define $f_i(x_i) = -\frac{\partial \phi}{\partial x_i}$, and we find the $x^* = (x_1^*, \ldots, x_n^*)$ that minimizes $\max_i f_i(x_i)$. By Theorem 2, the resulting point x^* has the property that $\frac{\partial \phi}{\partial x_i} = \frac{\partial \phi}{\partial x_j}$ whenever both x_i and x_j are positive, and $\frac{\partial \phi}{\partial x_i} \leq \frac{\partial \phi}{\partial x_j}$ when $x_i > 0$ and $x_j = 0$. (Recall that the partial derivatives are all negative, so f_i in our application of the abstract formulation is the negative of the corresponding partial derivative.) This implies that x^* is a local minimum for ϕ and thus, by strict convexity, it is the unique global minimum.

By using the priority algorithm and Theorem 2, we see that the choice of agents to assist with subsidies in the min-sum case proceeds via a priority rule, but one that is based neither on income nor on ruin probability. Rather, the priority given to agents is based on $\frac{\partial \phi}{\partial x_i}$ evaluated at 0

$$f_i(0) = \frac{w_i \beta_i \mu_i}{c_i^2}.$$

Interestingly, since the ruin probability is $\frac{\beta_i \mu_i}{c_i}$, it follows that the priority is, in fact, the product of three terms: the agent's weight, their ruin probability, and the reciprocal of their income.

4.2.3 Contrasting Prioritizations and Efficiency Loss

For both the min-max and min-sum objectives, we have seen a way to optimally allocate income subsidies in settings where agents have no initial wealth. Note that these results hold for any general distribution from which the shocks may be drawn and only require the mean size of the distribution. Another key take-away is that these algorithms inherently propose a *priority ordering* of the agents by need for the given objective. We can therefore ask: how different can these orderings be from one another? What about the ordering that simply uses the agents' income?

The three priority orderings under consideration are:

(I) By income $(c_1, c_2, \ldots, c_n)$, where $c_i \leq c_j$ if and only if $i < j$.

(II) By ruin probability $(\psi_1, \psi_2, \ldots, \psi_n)$, where $\psi_i \geq \psi_j$ if and only if $i < j$ used for the min-max objective.

(III) By the priority ordering $\frac{\partial \phi}{\partial x_i}$ used for the min-sum objective.

We investigate how different priority orderings (I), (II), and (III) can be from one another. Since the weights w_i used in the min-sum objective are a potential source of difference in these orderings for reasons independent of any of the other parameters, it is most interesting for the purposes of this question to consider the case in which these weights do not play a role; thus, we assume for this discussion that we have *unit weights*, with $w_i = 1$ for all agents i.

Lemma 3. *For any pair of orderings given by (I) income, (II) ruin probability, or (III) our solution given by Equation 4.2, there exist instances on which they are reverses of one another.*

Proof. For notational convenience, we first set $r_i = \beta_i\mu_i$. This value corresponds to the total amount of shock experienced per unit time. Therefore, the probability of ruin is $\psi_i(c_i, 0, \beta_i, \mu_i) = r_i/c_i$ and

$$\frac{\partial\phi}{\partial x_i} = \frac{-r_i}{(c_i + x_i)^2}.$$

We assume that we have agents with parameters:

$$((c_1, r_2), (c_2, r_2), \ldots, (c_n, r_n)).$$

Consider a set of n agents labeled $i = (1, 2, \ldots, n)$ with income $c_i = 1 + i\epsilon$ for some $\epsilon > 0$. The shock volume r_i is chosen to equal $(0.5 + 4i\epsilon)^2$. The priority ordering by income in this setting is $(1, 2, \ldots, n)$. On the other hand, the priority ordering by $\frac{\partial\phi}{\partial x_i}$ would rank the agents in increasing order of

$$-\frac{r_i}{c_i^2} = -\frac{(0.5 + 4i\epsilon)^2}{(1 + i\epsilon)^2}.$$

This quantity decreases as i increases. Therefore, the priority ordering given by the optimal solution for the min-sum objective is the reverse of the priority ordering by the agents' income.

This example also gives us a case where the priority ordering by income is the reverse of the priority ordering by ruin probability. In particular, the ruin probability is given by

$$\frac{r_i}{c_i} = \frac{(0.5 + 4i\epsilon)^2}{(1 + i\epsilon)}.$$

This value increases as i increases.

Finally, to note that the priority orderings given by (II) the ruin probability and (III) the optimal solution for the min-sum objective can the reverse of one another, we construct an example as follows: keep r_i to be the same as above, but let $c_i = 1 + 4i\epsilon$. Then, the probability of ruin will be

$$\frac{r_i}{c_i} = \frac{(0.5 + 4i\epsilon)^2}{(1 + 4i\epsilon)},$$

which is, again, increasing in i. On the other hand, the $\frac{\partial \phi}{\partial x_i}$ values are

$$-\frac{r_i}{c_i^2} = -\frac{(0.5 + 4i\epsilon)^2}{(1 + 4i\epsilon)^2},$$

which decreases as i increases. Therefore, the priority orderings given by the ruin probability and the optimal solution for the min-sum objective can be reverses of one another. □

On the other hand, there are some dependencies among the priority orderings. One natural one follows by observing that if reciprocal income and probability of ruin determine the same priority ordering, then since the priority for the min-sum objective $\frac{\partial \phi}{\partial x_i}$ is the product of these two numbers, it too must follow the same ordering. (Note, again, that we assume unit weights here.)

Corollary 4. If the priority orderings by income and the ruin probabilities are the same, then the priority ordering given by the optimal solution for the min-sum objective (with unit weights) will also be the same.

These above examples show that the order in which we assist families are impacted by:

- the input that we take into account: i.e., do we only use agents' income or do we also incorporate information about shocks?
- the objective function: i.e., are we optimizing for a min-sum or a min-max objective?

These can have drastically different policy implications, which we will build on in the following section.

Given these observations, it is natural to ask what the potential cost of using a more naive prioritization scheme would be, and we find that it can be high. Namely, we find that there can be a gap of $\Omega(\sqrt{n})$ comparing our solution with both income and ruin probability.

Example 5. Suppose we have three types of agents A, B, and C, and all agents have $\beta_i = 1, u_i = 0$. We describe each of the three types of agents below:

A: Each agent has $\mu_i = m$ and $c = \frac{11}{10}m$. There are m agents of this type.

B: This is one agent with $\mu = m^3$ and $c = m^3$.

C: Each agent has $\mu_i = 0$ and $c_i = 1$. There are m^2 agents of this type.

The ruin probabilities for agents of type A, B, and C are $10/11, 1$, and 0, respectively, and there are a total of $m^2 + m + 1$ agents. The expected number of agents experiencing ruin is $\frac{10m}{11} + 1$.

Suppose the designer has a budget of $\frac{1}{10}m^3$. We consider the three types of priority orderings:

I: A priority ordering by income would focus on type C agents. We divide up our budget evenly, and each agent would receive an income subsidy of $m/10$. Since these agents already had ruin probability 0, the expected number of agents experiencing ruin remains unchanged.

II: Next, we consider prioritizing the agents by ruin probability. In this case, we would give our entire budget to the agent of type B, reducing their ruin probability to $10/11$. In this case, the expected number of agents experiencing ruin is still $\Theta(m)$.

III: Finally, we consider our solution: we would prioritize agents of type A, dividing up our budget evenly among the agents. In this case, the ruin probability of each agent in A is reduced to $O(1/m)$, and so the expected number of agents experiencing ruin is $O(1)$.

We therefore have a gap of $\Omega(m)$ between our solution and both the priority orderings that use income or ruin probability. Since there are $n = \Theta(m^2)$ agents, the gap is proportional to the square root of the number of agents.

4.3 Non-zero Wealth: The Exponential Case

The case of non-zero wealth becomes much more complex in general; for example, we do not have closed-form expressions for the ruin probability with general distributions as we do for the case of zero wealth. To get a sense for the properties of non-zero wealth in a setting that has complex behavior but still exhibits closed-form solutions, we consider the case when shocks are drawn from an exponential distribution F_i: the probability that a shock has size exceeding x is given by $e^{-\delta x}$ for a parameter $\delta > 0$, resulting in a mean shock size of $1/\delta$. By studying this special case, we highlight counter-intuitive results showing non-monotonicity of agent's welfare in the various parameters of interest as well as rich examples where priority orderings by income, wealth, and our solutions for optimal income and wealth subsidy can vary substantially.[3]

Specifically, suppose an agent has income c_i, initial wealth u_i, and experiences shocks at rate β_i with a size distribution that is exponential with mean $1/\delta_i$; then a result from the theory of ruin processes (Asmussen and Albrecher, 2010) shows that the ruin probability is

$$\psi(c_i, u_i, \beta_i, F_i) = \frac{\beta_i}{c_i \delta_i} e^{\left(\frac{\beta_i}{c_i} - \delta_i\right) u_i}. \tag{4.3}$$

Since the distribution F_i is fully characterized by the parameter δ_i, we will also write this as $\psi(c_i, u_i, \beta_i, \delta_i)$.

We consider both wealth and income subsidies, as well as subsidies that can combine contributions to both wealth and income. A wealth subsidy of z_i reduces the ruin probability from $\psi(c_i, u_i, \beta_i, \delta_i)$ to $\psi(c_i, u_i + z_i, \beta_i, \delta_i)$. For the min-max objective, we can proceed exactly as in the previous section, applying the priority algorithm from Section 4.2.1 using the functions $f_i(z_i) = \psi(c_i, u_i + z_i, \beta_i, \delta_i)$, and hence ordering agents by their ruin probabilities. For the min-sum objective, we can take the partial derivatives to define a priority ordering for the algorithm according to the framework of Section 4.2.1. The solution for the income subsidy proceeds in a similar fashion: again, using the ruin probabilities for the min-max objective and the partial derivatives of $\psi(c_i + x_i, u_i, \beta_i, \mu_i)$ for the min-sum objective.

We finally consider the case where the planner can give a mix of income and wealth subsidies. We assume that each unit of income subsidy counts toward

[3]Missing proofs and further details in this section can be found in Appendix A.1.

the budget at a factor of k times the contribution of each unit of wealth subsidy. So, we have a total budget of $B_u + kB_c = B$, where B_c is the total income subsidy and B_u is the total wealth subsidy allocated. In the appendix, we show that the optimal solution here directly applies the above solutions for income and wealth subsidy. Note that although a wealth subsidy is implemented as a one-time allotment while an income subsidy is paid out over time, it is natural to convert between them if we take the standard interpretation of income subsidies as coming from a fixed endowment that grows geometrically to support a fixed rate of payout.

4.3.1 Properties of Subsidies

The solutions outlined above and in the corresponding appendix section provide a number of different priority orderings, depending on the objective function and the nature of the subsidy (e.g., income subsidy or wealth wealth). Since these orderings impact which agents will receive subsidies first, it is natural to ask what the relationships among these priority orderings might be. For similar reasons as in the case with no initial wealth, we again set the weights w_i to be 1. The discussions to establish these results below as well as discussions around monotonicity can be found in Appendix A.1.

The min-max objective orders the agents by the probability of ruin. We first show that this ordering can be the reverse of the ordering given by income or wealth.

Lemma 6. *The priority ordering given by ruin probability (and likewise the wealth subsidy solution) can be the reverse of the priority ordering given by income and wealth,* *even when the priority ordering by income and wealth agree.*

Using the case when the agents' wealth is zero, we have shown that the priority ordering by ruin probability can be the reverse of the ordering given by the income subsidy solution. We can show a similar inversion with the wealth subsidy and the ruin probability using the case where shocks are drawn from an exponential distribution. In fact, we find that a stronger result holds.

Lemma 7. *The priority ordering given by the ruin probability can be the reverse of the priority ordering given by the optimal solution for income subsidy and wealth subsidy, even when the latter two coincide.*

A distinct question, and one whose answer is not a priori clear, is whether the priority orderings for income and wealth subsidies must be the same. We see that they are, in fact, not, indicating that which agents we prioritize depends on the kind of subsidy that is being allocated by the planner.

Lemma 8. *The priority ordering given by the optimal solution for income subsidy can be the reverse of the priority ordering given by the optimal solution for the wealth subsidy.*

This above result adds a key observation to the previous section. Namely, the order in which we assist families is impacted not only by the parameters that we take into account and the objective function but also by the type of intervention; a family that may be last in line to receive assistance under one intervention may be first under another.[4]

[4] These observations, we believe, are of independent interest and we discuss some of the implications in the discussion section as well as the Part IV on the roles for computing in social change.

Monotonicity Properties. These contrasting prioritizations surface a natural question about whether the quantities we are considering — ruin probability, optimal income subsidy value, optimal wealth subsidy value — are monotone in an agent's parameters. It is easy to note that the ruin probability exhibits monotonicity in all the parameters c_i, u_i, β_i, and δ_i. That is, holding all other variables constant, if an agent's income or wealth is lower, or they experience more frequent or larger shocks, then the ruin probability increases. However, we find that this is not the case for the optimal income and wealth subsidy values, which are only monotone in (complementary) subsets of the parameters. The table below contains a summary of these results, where "yes" indicates that the priority ordering given by the row is monotone in the parameter given by the column and "no" indicates that it is not. Recall again that the priority given by the ruin probability is one given by the optimal solution for income or wealth subsidies for the min-max objective.

Type	c	u	β	δ
Wealth subsidy: min-sum	no	yes	no	no
Income subsidy: min-sum	yes	no	yes	yes
Ruin probability	yes	yes	yes	yes

Table 4.1: Monotonicity results for income and wealth subsidies for min-max and min-sum by each of the four parameters: $c_i, u_i, \beta_i, \delta_i$. Entries "yes" indicate that the subsidy is monotone in the corresponding parameter and "no" indicates that it is not. Both the income and wealth subsidies use the ruin probability as the agents' priority ordering for the optimal solution for the min-max objective.

Therefore, for instance, a family i may go up in our priority ordering if we fix u_i, β_i, and δ_i, but their income increases for the min-sum objective where the designer is providing a wealth subsidy. This, however, is not possible for the min-sum objective under income subsidy or for the min-max objective under either subsidy scheme.

4.4 Non-zero Wealth: General Distributions

The previous sections characterize optimal subsidies when all agents have positive drift, together with either zero initial wealth or shock distributions satisfying a specific functional form. We now consider the case of general shock distributions, arbitrary initial wealth, and arbitrary drift. This section contains three results:

- Lemma 9: a polynomial-time algorithm for the min-max objective when agents have arbitrary shock distributions, initial wealth, and drift.
- Theorem 10 (main result): a Fully Polynomial-Time Approximation Scheme (FPTAS) for the min-sum objective when agents have arbitrary shock distributions, initial wealth, and drift.
- Proposition 11: a proof that the min-sum objective is (weakly) NP-hard in general, implying that we should not expect better than an FPTAS without further assumptions on the problem.

4.4.1 Min-Max via Binary Search

We first show how to allocate income subsidies so as to optimize the min-max objective. Intuitively, our algorithm looks like a binary search for the optimal min-max value X. For each potential guess p of X, we just need to see, for each agent, how much income subsidy is required to achieve a ruin probability p. If the sum of these subsidies exceeds B, then it is infeasible to have min-max value p. Otherwise, it is feasible. So, we can repeatedly guess p in binary search and converge quickly to the optimum. Lemma 9 essentially formalizes this intuition while being careful about the cost of certain operations. For example, throughout this section, we will assume for simplicity of exposition that $\psi^{-1}_{u,\beta,F}(\cdot)$ can be computed in $O(1)$ operations, where $\psi_{u,\beta,F}(c) = \psi(c, u, \beta, F)$ (that is, the minimum x such that $\psi(x, u, \beta, F) \leq p$ can be computed in $O(1)$ operations for all p). We briefly discuss at the end of this section ways in which this assumption can be relaxed.

Below, we seek to allocate income subsidies $x_1, \ldots, x_n$ with $\sum_i x_i = B$ such that agent i receives subsidy x_i in a way that minimizes $\max_i\{\psi(c_i + x_i, u_i, \beta_i, F_i)\}$.

Lemma 9. *Let X denote the optimum value for the min-max objective for any given instance. Then, for any $\delta > 0$, a solution with min-max value $X + \delta$ can be found in time* polynomial$(n, \log(1/\delta))$.

Proof. The algorithm is based on binary search. We first need a subroutine to check, for a given value p, whether it is feasible to subsidize all agents to probability of ruin at most p. To this end, simply compute $\psi^{-1}_{u_i,\beta_i,F_i}(p)$ for all i, and update $x_i := \max\{0, \psi^{-1}_{u_i,\beta_i,F_i}(p) - c_i\}$. If $\sum_i x_i \leq B$, then these choices of x_i explicitly witness that it is feasible to have min-max objective $\leq p$. If not, then they explicitly prove that with a budget constraint of B, some agent must have ruin probability exceeding p.

It therefore suffices to do binary search on the interval $[0, 1]$. Each iteration of binary search takes $O(n)$ operations, and therefore doing $\log(1/\delta)$ iterations of binary search takes $O(n \log(1/\delta))$ operations. After $\log(1/\delta)$ iterations, we will have a window $[X, X + \delta]$ where we explicitly found a choice of subsidies guaranteeing min-max objective $X + \delta$, and also proved that better than X is not possible. Thus, we have an additive δ approximation in the desired time. $\square$

4.4.2 Min-Sum via Knapsack

We now show how to make use of our min-max approximation as a subroutine to provide an FPTAS for the min-sum objective. Our approach barely uses any structure of the $\psi_{u_i,\beta_i,F_i}(\cdot)$ function, and is essentially an FPTAS to minimize $\sum_i f_i(x_i)$ subject to $\sum_i x_i \leq B$ for any no-increasing functions $f_i(\cdot)$. Our algorithm is also very similar to the FPTAS for Knapsack via dynamic programming.

Theorem 10. *Let X denote the optimum value for the min-sum objective for any given instance. Then, for any $\varepsilon, \delta > 0$, a solution with min-sum value $(1+\varepsilon)X + \delta$ can be found in time* polynomial$(n, 1/\varepsilon, \log(1/\delta))$.

Proof. The algorithm proceeds in two phases. First, we need to figure out the scale of the optimum solution. Then, we achieve an additive approximation at the appropriate scale by adapting the FPTAS for the weighted knapsack problem (Vazirani, 2013).

For phase one, we simply solve the min-max objective using Lemma 9 up to accuracy $\delta/(2n)$, in time polynomial$(n, \log(n/\delta))$. Let C denote the value of the solution output by the min-max subroutine. If $C \leq \delta/n$, then certainly the min-sum objective for this same solution is at most δ, which satisfies the desired guarantee (as $X \geq 0$). Otherwise, we know from Lemma 9 that it is certainly not possible to get the ruin probabilities to sum to less than $C - \delta/(2n)$ (since, otherwise, the max would be at most $C - \delta/(2n)$ as well, which would contradict Lemma 9). As $C \geq \delta/n$, this means that the optimum is at least $C/2$. Therefore, if we set $\eta := \varepsilon C/(2n)$ and get an additive $n\eta$ approximation, this will be a multiplicative $(1+\varepsilon)$ approximation.

Observe also that our min-max subroutine explicitly outputs a solution with min-sum value at most Cn (because the maximum ruin probability is at C). We thus know that the optimum lies somewhere between $C/2$ and Cn. These bounds are sufficient to use a dynamic program similar to that for Knapsack.

To create our Dynamic Program, first define:

$$f_i(x_i) := \lceil \psi(c_i + x_i, u_i, \beta_i, F_i)/\eta \rceil \cdot \eta.$$

That is, $f_i(x_i)$ is $\psi(c_i + x_i, u_i, \beta_i, F_i)$ rounded up to the nearest integer multiple of η. Note that this only makes any potential solution worse by at most an additive factor of η. After this rounding, we claim we can find the optimal solution with dynamic programming. Briefly observe that $f_i^{-1}(\cdot)$ can be computed with one black-box call to $\psi^{-1}_{u_i,\beta_i,F_i}(\cdot)$ (which are still assuming can be computed in $O(1)$ operations for ease of exposition).

Let $G(j,k)$ denote the minimum possible budget that suffices to guarantee $\sum_{i \le j} f_i(x_i) = k \cdot \eta$. We claim this can be found using the recurrence below.

$$G(j,k) = \min_{\ell}\{G(j-1, k-\ell) + f_i^{-1}(\ell\eta)\},$$

where ℓ ranges in $\{0, 1, \ldots, k\}$.

Observe first that the range of ℓ suffices, as all values are integer multiples of η. Observe also that the optimal solution for $G(j,k)$ must obtain some value $\ell\eta$ for $f_j(x_j)$, which means it must get $(k-\ell)\eta$ from the first $j-1$, and that the optimization for the first $j-1$ is exactly covered by $G(j-1, k-\ell)$. Finally, observe that each step requires taking a minimum over at most k terms, and that the entire DP table has $n * K$ terms, if we let k range in $\{0, \ldots, K-1\}$. The entire table can thus be filled in time polynomial(n, K). We will set $K := n^3/\varepsilon$.

The solution we seek is the minimum k such that $G(n,k) \le B$. Observe that as $K := n^3/\varepsilon$, it is certainly the case that $G(n,K) \le B$, as this corresponds to a solution of value $n^3 \cdot \eta = Cn^2$, which is guaranteed to exist by our work in phase one. Observe that this dynamic program finds an optimal allocation, up to an additive $n\eta = \varepsilon C/2$. So, as the optimum is at least $C/2$, this is at most $(1+\varepsilon)\text{OPT}$, as desired. $\square$

4.4.3 NP-Hardness and Further Computational Considerations

In this section, we prove an NP-hardness result showing that one should not expect an exact solution in polynomial time without some assumptions, and also discuss relaxations of the assumption that $\psi^{-1}_{u,\beta,F}(\cdot)$ can be computed in $O(1)$ operations.

First, recall that the approach used in Theorem 10 is quite general: it essentially provides an FPTAS for minimizing $\sum_i f_i(x_i)$ subject to $\sum_i x_i \leq B$ for *any* non-increasing functions such that $f_i^{-1}(\cdot)$ can be computed efficiently. We first show that this problem is NP-hard in general, even for fairly simple functional forms of $f_i(\cdot)$.

Proposition 11. Let MIN-SUM take as input explicit descriptions of n functions $f_i(\cdot)$, and output $\min_{\vec{x}, x_i \geq 0 \forall i, \sum_i x_i \leq B}\{\sum_i f_i(x_i)\}$. Then, MIN-SUM is (weakly) NP-hard, even when each $f_i(\cdot)$ takes the form of $\min\{1, g_i(\cdot)\}$, where $g_i(\cdot)$ is convex.[5]

Proof. We reduce from SUBSET-SUM. Given an instance of SUBSET-SUM with $A_1, \ldots, A_n$ for which we want to know whether there is a subset that sums to exactly C, we do the following:

[5]Observe that without the minimum with 1, this is just convex minimization, which can be solved in polynomial time.

- Let $A \geq \max_i A_i$.
- Define $g_i(x_i) := \max\{1 - A_i/A, 1 + (A_i - 2x_i)/A\}$. Observe that $g_i(\cdot)$ is the maximum of linear functions and therefore convex. Now define $f_i(x_i) := \min\{1, g_i(x_i)\}$. Note that $f_i(0) = 1$.
- Input $f_1(\cdot), \ldots, f_n(\cdot), C$ to MIN-SUM. If there is a solution with value $n - C/A$, answer "yes" to SUBSET-SUM. Otherwise, answer "no."

The above reduction clearly runs in polynomial time. Therefore, it only remains to show whether it is correct. First, observe that the derivative of $f_i(\cdot)$ is either 0 or $-2/A$ everywhere. In particular, $f_i'(x_i) = 0$ on $[0, A_i/2]$, $f_i'(x_i) = -2/A$ on $[A_i/2, A_i]$, and $f_i'(x_i) = 0$ on $[A_i, \infty)$.

Therefore, if we look at a candidate solution $\vec{x}$, we necessarily have $f_i(x_i) \geq 1 - x_i/A$. This is simply because the average derivative on $[0, x_i]$ is at least $-1/A$ for all x_i. The inequality is tight only at $x_i = A_i$. From this, we conclude that if there is a subset such that $\sum_{i \in S} A_i = C$, we could consider setting $x_i = A_i$ for all $i \in S$ and $x_i = 0$ otherwise. This would yield a solution with value $n - C/A$. This is because all $i \in S$ will have $f_i(x_i) = 1 - A_i/A$, and so our total solution will have value $n - C/A$.

Similarly, if there exists a solution with value $n - C/A$, this implies that the gain in value for every unit of budget is exactly $-1/A$ (because no budget contributes better than $-1/A$, and our total budget is B). For each agent, this implies that either they consume no budget ($x_i = 0$) or they consume exactly A_i of budget. Because the entire budget is consumed at bang-for-buck of $-1/A$, if S denotes the agents who receive non-zero subsidy we must have $\sum_{i \in S} A_i = C$. Therefore, the reduction is correct. □

It is certainly possible that for certain functional forms of F_i (e.g., exponentially distributed), that the min-sum objective can be optimized exactly in polynomial time. But the above hardness shows that and FPTAS is likely the best one should hope for without further assumptions. Finally, we briefly discuss further computational considerations with regards to computing $\psi^{-1}_{u_i,\beta_i,F_i}(\cdot)$.

Lemma 12. *Let $f(\cdot)$ be non-increasing, and have a domain of $[0, B]$. Then if $f(\cdot)$ can be computed in $O(1)$ operations, for all x, a y satisfying $y \leq f^{-1}(x) \leq y+\delta$ can be computed in $O(\ln(B/\delta))$ operations.*

Proof. Let the input query be x. If $f(B) > x$, then the inverse is undefined. If $f(0) < x$, then the inverse is also undefined. Otherwise, an inverse exists in $[0, B]$. From here, we proceed with binary search. After $\log_2(B/\delta)$ steps, we will have a window of the form $[y, y + \delta]$ where $f^{-1}(x)$ certainly lies in this window. So output this y. □

This lemma alone does not transparently affect the approximation guarantees — the issue is that perhaps a little error in $y \approx f^{-1}(x)$ may cause $f(y) \gg x$. With one extra assumption on f, however (a Lipschitz condition), this is useful. Below, we say that a function $f(\cdot)$ is (L, δ)-Lipschitz if for all x, y with $|x-y| \leq \delta/L$, $|f(x) - f(y)| \leq \delta$.

Corollary 13. Let $\psi_{u_i,\beta_i,F_i}(\cdot)$ be (L, δ)-Lipschitz. Then if $\psi_{u_i,\beta_i,F_i}(\cdot)$ can be computed in $O(1)$ operations, an additive $O(\delta)$-approximation to the min-max objective can be found in time polynomial$(n, \log(LB/\delta))$, and a multiplicative $(1 + \varepsilon)$, additive δ approximation to the min-sum objective can be found in time polynomial$(n, 1/\varepsilon, \log(LB/\delta))$.

Proof Sketch. Using Lemma 12, a y satisfying $y \leq \psi^{-1}_{u_i,\beta_i,F_i}(x) \leq y + \delta/L$ can be found in $O(\ln(LB/\delta))$ operations. Therefore, this value of y also satisfies $f(y) \leq x + \delta$, as ψ_{u_i,β_i,F_i} is (L, δ)-Lipschitz. Using this instead of a true inverse oracle accumulates an additional additive error of δ in both proofs, which can be accommodated into the existing guarantees. □

Intuitively, the need for *some* Lipschitz condition on $\psi_{u_i,\beta_i,F_i}(\cdot)$ (if we only wish to have black-box access to $\psi_{u_i,\beta_i,F_i}(\cdot)$ and not $\psi^{-1}_{u_i,\beta_i,F_i}(\cdot)$) is because if ψ_{u_i,β_i,F_i} jumps instantaneously from (say) 1 to 0, it is impossible to detect exactly where. Thus, in order to ensure we are competitive with the optimum, we would also need to violate the budget constraint by a little bit. Even without any Lipschitz condition, this approach suffices:

Lemma 14. *If all $\psi_{u_i,\beta_i,F_i}(\cdot)$ can be computed in $O(1)$ operations. Then if X denotes the optimal solution to the min-max objective with budget B, a solution with min-max value $X + \delta$ using budget $B + \delta$ can be found in time* polynomial$(n, \log(B/\delta))$. *If Y denotes the optimal solution to the min-sum objective with budget B, a solution with min-sum value $(1+\varepsilon) \cdot Y + \delta$ using budget $B + \delta$ can be found in time* polynomial$(n, 1/\varepsilon, \log(B/\delta))$.

Proof Sketch. Similar to Corollary 13, the proof follows by observing that a $y \leq \psi^{-1}_{u_i,\beta_i,F_i}(x) \leq y + \delta/n$ can be found in $O(\ln(nB/\delta))$ operations. Using $y + \delta/n$ as $\psi^{-1}_{u_i,\beta_i,F_i}(x)$ results in no additional error in ruin probabilities, but does cost at most an additional δ/n in budget consumed. Chasing through the rest of the previous proofs, the same guarantees are achieved, but after consuming an additional δ budget. □

4.5 Discussion

In this chapter, we propose a model of economic welfare that incorporates agents' income and initial wealth as well as income shocks, and we analyze the model using results from the theory of ruin processes. We consider a problem faced by a planner who would either like to maximize the welfare of the most vulnerable agent or minimize the number of agents that experience ruin. Our analysis reveals several insights into the role of income shocks on economic welfare. For instance, we find that agents may appear to be less vulnerable when considering welfare measures that simply use income than measures we introduce in this chapter. And, in fact, we also find that even measures proposed in this chapter — i.e., ruin probability or those given by our solutions for income or wealth subsidies — can be drastically different from one another. Therefore, care must be taken not only to ensure that variables that impact individuals' economic well-being (such as income shocks) are taken into account, but also in picking the desired objective as well as the type of subsidy to be given out. These decisions are best approached as an iterative conversation and collaboration with domain experts and the setting in which insights from such optimization-based approaches may be deployed.

The different forms of subsidies that we consider in this chapter closely resemble different assistance programs or proposed subsidies. Income subsidies are reflected in programs such as the Supplemental Nutrition Assistance Program or housing vouchers which reduce families' monthly expenses, thereby leaving more reserves for families month-to-month. (In our model, this roughly corresponds to adding an income subsidy x_i to c_i.) Wealth subsidies may resemble proposed policies such as "baby bonds" or inheritance taxation to alleviate racial wealth inequalities (Hamilton and Darity Jr, 2010; Shapiro et al., 2004)

Our Model. There are various assumptions about our model that warrant further discussion and also leave open directions for further work. We treat income as a steady stream. We incorporate disruptions to one's income as a form of shock. Cases in which an individual loses their job entirely, or experiences some other shock that essentially amounts to a long-term state change, may benefit from a different model that more directly incorporates such discrete transitions.

We also assume that income and initial wealths are known, well-defined values. In practice, it may be difficult to quantify these, and especially reserves. For instance, reserves such as savings may be simple to represent, whereas other forms of reserves such as social capital or having "outside options" such as ability to rely on family for certain types of shocks may be harder to quantify. Note that we address some of this via the weighted version of our problem, with potentially different weights assigned to each agent, but there are many options in how best to determine these weights and the weights themselves may also be dynamic. Similarly, we assume that the agents' incomes are fixed and they do not make choices in consumption. In many settings, agents may adjust their consumption or make investments in response to such interventions.

We further assume time-independence. That is, given a fixed reserve and income flow, we assume that experiencing a shock at time t_1 versus t_2 has the same impact. However, shocks may be sensitive to the time when they are experienced, if the impact of the shock has the potential to be felt longer in some cases than others. (For example, consider certain health outcomes, or interactions with the criminal justice system at different ages or in different circumstances.) Our model implicitly assumes that our process is running during a time when shocks, weights, threshold for ruin and other such parameters are time-independent.

Societal Implications. There are a number of further societal considerations inherent in the problem we are studying. One such consideration is connected to the assumption that shocks can be observed or their distributions can be known in some instances. As discussed in the introduction, obtaining this information can be intrusive into the lives of vulnerable individuals. While some information, such as interactions with the criminal justice system, may already be data that a planner — such as a government assistance program — already has access to, others — such as the dissolution of a romantic or other close personal relationship — may represent private and sensitive information. There is a rich line of work exploring the class differentials in privacy loss and privacy violations, especially for data gathered as part of government assistance programs (Gilman, 2011; Marwick et al., 2017; Peppet, 2011). This suggests a number of open questions, including how to allocate subsidies effectively when we only have noisy data about agents — including cases where the noise comes from the planner deliberately limiting their data collection on shocks.

A separate issue is concerned with setting the objective function. In this chapter, we study the min-max and min-sum objectives, which aim to maximize the welfare of the agent who is most susceptible to experiencing ruin and minimizing the expected number of agents that may experience ruin. Both of these are well-studied objectives in the literature and also correspond to objectives that are motivated in practice. For instance, in the motivating literature related to eviction, a planner might aim to minimize the maximum likelihood that any family will get evicted or the planner may wish to minimize the expected number of families who may experience eviction. Social programs that work by identifying and targeting families who may benefit most from such programs are common in various domains, including low-income housing assistance programs and poverty-reduction efforts more generally.

At the same time, social policy may also provide such provisions to all families. There is rich work across the social sciences and public policy exploring the role of targeting in poverty reduction ranging from broader survey (Akerlof, 1978; Elster, 1992; Mirrlees, 1971), to the role of self-targeting and other mechanisms that may drive the process (Alatas et al., 2016; Nichols and Zeckhauser, 1982), to work focused on specific domains such as medicine (Persad et al., 2009). Note that in some domains, there may be a choice between policy that takes a "targeting" versus "universalist" approach, where programs would build a universal floor for all families (Mkandawire, 2005; Skocpol, 1991).

Recent work including by Eubanks has highlighted the complex interaction between public services and algorithmic decision-making tools as well as challenges that arise in implementing such solutions in real-world settings (Eubanks, 2018b,a). Prior to this work, researchers and practitioners in economics and computation have studied the societal and ethical considerations that emerge when allocating scarce societal resources and the role of repugnant markets (Calabresi and Bobbitt, 1984; Elster, 1992; Roth, 2007, 2008). For further discussions on ethical and societal considerations of algorithmic and mechanism design work in settings such as the ones considered in this work, see Part V.

CHAPTER 5

A TRUTHFUL MECHANISM FOR ONE-SIDED MATCHING

Resource allocation in public sector domains, whether it is the allocation of subsidies to families, assignment of students to public schools, or matching of doctors to hospitals, often surface fundamental mathematical questions. One such mathematical problem is what is known as the *house allocation* problem of Hylland and Zeckhauser (1979): in this problem, a set of agents n agents are to be matched, one-to-one, to a set of n items, such as houses or seats in public schools. This problem is further characterized by the assumptions that there are no monetary transfers and agents can have heterogeneous preferences over the items. For instance, in assigning students to public schools or assignment of low-income housing resources to families, families may have preferences for locations in specific neighborhoods closer to their communities or for schools or houses that meet specific needs that they might have. And, they might strategically report their preferences in an attempt to secure better resources for themselves.

More concretely, we assume that each agent i has a value $v_{i,j}$ for each item j. A randomized matching mechanism outputs a probability distribution over matchings, which corresponds to a doubly-stochastic matrix p, providing the probability $p_{i,j}$ that i will be matched to j; the expected utility of i in p is $\sum_j v_{i,j} p_{i,j}$. The goal in this setting is to generate a fair and efficient randomized matching, which crucially depends on the values of the agents, and the main obstacle is the fact that the $v_{i,j}$ values of each agent i are private information of this agent. Therefore, a successful mechanism needs to elicit the agents' preferences and output a desired matching; on the other hand, each agent's goal is to maximize her expected utility, so an agent can strategically misreport her preferences if this increases her expected utility.

This tension between the objectives of the designer and those of the participants lies at the core of the sub-field of economics and computation known as *mechanism design*. Most of the proposed solutions in the mechanism design literature, however, leverage monetary payments as the main tool that the designer can use to incentivize truthful reporting by the agents. This, however, may be undesirable or infeasible in public sector settings. [1] This requirement makes the mechanism design problem considered in this chapter particularly demanding. In the absence of monetary payments, an alternative tool for simulating the impact of payments is to use "money burning" (Hartline and Roughgarden, 2008). In our setting, this could correspond to keeping some of the items unmatched with positive probability, thus penalizing the agents just like monetary payments would. But, in many settings, including the house allocation problem studied in this chapter, this would be unacceptable (e.g., it would imply that some agents may remain homeless while some houses remain unoccupied).

Our main result in this chapter is a novel application of random sampling that enables the use of known money burning mechanisms, while ensuring that every agent and item is matched. In other words, our technique takes advantage of the improved incentives that these money-burning mechanisms provide, but without suffering their most important drawback. To verify the usefulness of this approach, we combine it with the *partial allocation* (PA) mechanism of Cole et al. (2013), giving rise to a new mechanism that incentivizes the agents to always truthfully report their cardinal preferences, i.e., their $v_{i,j}$ values, and

[1] For instance, we would like to assign families to low-income housing resources without any monetary transfers.

yields an outcome p that is approximately both fair and efficient.[2] We measure the performance of our mechanism using the canonical benchmark defined by the Nash bargaining solution and show that our mechanism outperforms the standard mechanisms with the same, or weaker, incentive properties.

The literature on one-sided matching has considered three main approaches, none of which gives rise to mechanisms that are both truthful and obtain a non-trivial approximation of the aforementioned benchmark. Hylland and Zeckhauser (1979) propose the *competitive equilibrium from equal incomes* (CEEI), which depends on the $v_{i,j}$ values in a non-trivial way, but it provides the agents with strong incentives to misreport these values, especially for small problem instances. The *random serial dictatorship* (RSD), or random priority, mechanism is an important mechanism with a long history in practice. This mechanism randomly orders the agents and, following this order, gives to each agent her favorite item among the ones that are still available. RSD is an *ordinal* mechanism: it requires only the ordinal preferences of each agent, i.e., only her ranking of the items from most to least preferred. It elicits this information truthfully, but its outcomes can be very inefficient. The *probabilistic serial* (PS) mechanism of Bogomolnaia and Moulin (2001) is another ordinal mechanism, and its outcome is computed by continuously allocating to each agent portions of her most preferred item that has not already been fully allocated. PS satisfies an ordinal notion of efficiency, but it achieves only a trivial approximation of our stronger benchmark, and it is not truthful. A more detailed discussion regarding these mechanisms and other related work is provided in Chapter 2.

Aiming to provide stronger efficiency and fairness guarantees compared to

[2] Zhou (1990) shows that there is no mechanism that is truthful, symmetric, and Pareto efficient; thus, some notion of approximation is necessary.

known mechanisms, we consider a cardinal benchmark: the well-studied *Nash bargaining solution*, proposed by Nash (1950). Given a *disagreement point*, i.e., the "status quo" that would arise if negotiations among the agents were to break down, the Nash bargaining solution is the outcome that maximizes the product of the agents' marginal utilities relative to their utility for the disagreement point. This outcome indicates the utility that each agent "deserves," so we use this utility as the benchmark for that agent. The choice of disagreement point can depend on the application at hand: if a buyer and a seller are negotiating a transaction, the disagreement point could be that the seller keeps the goods and the buyer keeps her money. In one-sided matching markets, the disagreement point needs to be a matching because leaving an agent without a house is infeasible. Since all agents have symmetric claims on the items when entering the market, we let the disagreement point be a matching chosen uniformly at random, ensuring that each agent is equally likely to be matched to each item. The Nash bargaining solution therefore corresponds to the doubly-stochastic matrix p that maximizes $\prod_i(\sum_j v_{i,j}p_{i,j} - o_i)$, where $o_i = \frac{1}{n}\sum_j v_{i,j}$ is the expected utility of agent i for an item chosen uniformly at random.

Since no truthful and symmetric mechanism can guarantee Pareto efficiency (Zhou, 1990), it is clearly impossible for a truthful mechanism to implement the Nash bargaining solution, which is symmetric and Pareto efficient. Thus, we consider the problem of approximating this solution. Specifically, the Nash bargaining solution defines the utility that each agent deserves and our goal is to ensure that every agent receives a good approximation of that benchmark. Formally, a mechanism is a β-approximation if the utility of each agent is at least a β fraction of her utility in the Nash bargaining solution. Note that, once the valuations of each agent i are adjusted by subtracting o_i, then our objective corresponds to the *Nash social welfare* (NSW), which has recently received a lot of attention in the fair division literature (e.g., Brânzei et al., 2017; Barman et al., 2018; Cole and Gkatzelis, 2018; Caragiannis et al., 2016; Garg et al., 2018). The NSW maximizing outcome is *proportionally fair* in that it satisfies a multiplicative version of Pareto efficiency, namely, the utility of an agent cannot be increased by a multiplicative factor without decreasing the product of utilities of other agents by a greater multiplicative factor.

En route to proving our mechanism's approximation bounds, we also provide an analysis of the Nash bargaining solution with respect to its *population monotonicity*, which is of independent interest. It has long been known that, unlike the Kalai-Smorodinsky solution, the Nash bargaining solution can violate population monotonicity for some instances of the bargaining problem (Thomson, 1983; Thomson and Lensberg, 1989). That is, there exist instances where removing some of the agents and computing the updated Nash bargaining solution can decrease the utility of some of the remaining agents. When allocating items among competing agents, this lack of monotonicity is somewhat counterintuitive. Why would the decreased competition from agents departing the market not lead to (weakly) increased utility for the agents remaining in the market? Indeed, we show that population monotonicity can be violated in the Nash bargaining solution for matching markets. Effectively, the constraint that the allocation is a distribution over perfect matching introduces positive externalities between agents.

In order to quantify the extent to which one of the remaining agents' utility can drop after such a change in the agent population, the bargaining literature in economics introduced the *opportunity structure* notion (e.g., see the book by

Thomson and Lensberg (1989), and references therein). This structure identifies the largest factor by which a remaining agent's utility can drop after some subset of agents is removed. In fact, resembling the standard computer science approach, the opportunity structure is defined as the *worst-case* factor over all instances, all removed subsets of agents, and all remaining agents. In this chapter, we provide essentially tight upper and lower bounds for this factor in the context of matching markets, showing that in carefully designed worst-case instances, this factor can grow faster than a polylogarithmic function of the number of agents, yet slower than any polynomial. Apart from the broader interest in understanding this measure in matching markets, we show that the upper bound on the population non-monotonicity provides, up to constant factors, an upper bound on the approximation factor of the truthful matching mechanism that we define.

Our Results

In this chapter, we introduce a random sampling technique which allows us to translate non-trivial truthful one-sided matching mechanisms that may produce *partial* matchings (i.e., possibly leaving some agents unmatched) into ones where

1. every agent is always assigned an item, and
2. the incentives for truthful reporting of preferences are maintained.

For example, the truthfulness guarantee of the PA mechanism of Cole et al. (2013) depends heavily on its ability to penalize the agents that cause inconvenience to others; it thereby ensures that none of these agents is misreporting their preferences. Since monetary payments are prohibited, this mechanism penalizes the agents by assigning positive probability to outcomes that leave them unmatched. Such partial matchings, however, are unacceptable in the house allocation problem. Every agent, no matter what values she reports, needs to be guaranteed an item, and this constraint significantly restricts our ability to introduce penalties. Nevertheless, we show that we can still recreate such penalties by using random sampling. Applying our sampling technique to the PA mechanism, we define the *randomized partial improvement* (RPI) mechanism, which significantly outperforms all the standard matching mechanisms with respect to the Nash bargaining benchmark.

In essence, the RPI mechanism endows agents with a baseline allocation given by a uniformly random item and then uses the PA mechanism to improve the agents' utility relative to this baseline. In reality, it is not possible to simultaneously maintain the baseline and offer improvements to all agents, so RPI circumvents this impossibility by imposing these two conditions on a sample of half of the agents instead. With half the agents (but all of the items) there is sufficient flexibility to faithfully implement the PA mechanism with the outside option of a uniform random house. After finalizing the allocation of the sampled agents, RPI then recursively allocates the unallocated portions of the items to the remaining agents.

As an intermediate step toward the theoretical analysis of RPI's approximation factor, we study the extent to which population monotonicity may be violated in a one-sided matching market instance. We refer to an instance as ρ-utility monotonic if removing a subset of its agents can *decrease* a remaining agent's utility in the new Nash bargaining solution by a factor of no more than ρ.

We show that, for a very carefully constructed (and somewhat contrived) family of instances, ρ can be as high as $\Omega(2^{\sqrt{\log n}/2})$ and we complement this bound with an essentially tight upper bound, by proving that for any one-sided matching instance ρ is no more than $O(2^{2\sqrt{\log n}}) \subseteq o(n^{\epsilon})$ for any constant $\epsilon > 0$.

Apart from the broader interest in understanding the extent to which the Nash bargaining solution may violate population monotonicity, our upper bound on ρ also directly implies an upper bound for the approximation factor of RPI. Specifically, we prove that RPI guarantees to *every* agent a $4\,e\,\rho$ approximation of the utility that she gets in the Nash bargaining benchmark. Therefore, as a corollary, we conclude that RPI approximates the Nash bargaining benchmark within $O(2^{2\sqrt{\log n}}) \subseteq o(n^{\epsilon})$ for any constant $\epsilon > 0$, even with the worst case choice of ρ. In stark contrast to this upper bound, which is strictly better than any polynomial, we show that the approximation factor of all ordinal mechanisms (even ones that are not truthful, such as probabilistic serial) grows linearly with the number of agents. Therefore, our mechanism significantly outperforms all ordinal mechanisms while at the same time satisfying truthfulness.

5.1 Preliminaries

Given a set N of n agents and a set M of n items, a randomized matching can be represented by a doubly-stochastic matrix p of marginal probabilities, where $p_{i,j}$ denotes the marginal probability that agent i is allocated item j. Clearly, any probability distribution over matchings implies a double-stochastic matrix, and the Birkhoff-von-Neumann theorem shows that *any* doubly-stochastic matrix can be implemented as a probability distribution over matchings. Denote by v a matrix of agent values where $v_{i,j}$ is the value of agent i for item j. The expected utility of agent i for random matching p is $u_i = \sum_{j \in M} v_{i,j}\, p_{i,j}$. The random matching p that a mechanism outputs when the agents' reported values are v is denoted by $p(v)$.

For each agent i, her values $v_i = (v_{i,1}, \ldots, v_{i,n})$ are private and a matching mechanism must be designed to properly elicit them. A mechanism is *truthful* if it is a dominant strategy for each agent i to report her true values. That is, if we let $p(w_i, v_{-i})$ denote the outcome of the mechanism when agent i reports values w_i and all the other agents report values v_{-i}, then a mechanism is truthful if for every agent i, any matrix of values v, and any misreports w_i

$$\sum_{j \in M} v_{i,j}\, p_{i,j}(v) \;\geq\; \sum_{j \in M} v_{i,j}\, p_{i,j}(w_i, v_{-i}).$$

Our benchmark, formally defined in the following section, uses the *Nash social welfare* (NSW) objective on appropriately adjusted agent valuations. The NSW maximizing outcome is known to provide a balance between fairness and efficiency by maximizing the geometric mean (or, equivalently, the product) of the agents' expected utilities, i.e., $\max_p \prod_i \left(\sum_j v_{i,j}\, p_{i,j}\right)$. The partial allocation mechanism from Cole et al. (2013) provides a truthful approximation of that outcome and can be easily adapted to randomized matchings by interpreting fractional allocations as probabilities.

Definition 15. The *partial allocation* (PA) mechanism on values v works as follows:

1. Compute the doubly-stochastic matrix $p^{\mathrm{NSW}}(v)$ that maximizes the Nash social welfare.
2. For each agent i, compute f_i as follows:

(a) Let u_k be agent k's utility in $p^{\mathrm{NSW}}(v)$.

(b) Let u'_k be agent k's utility in $p^{\mathrm{NSW}}(v_{-i})$, i.e., in the NSW maximizing allocation with agent i absent and all other agents restricted to one unit, i.e., $\sum_j p^{\mathrm{NSW}}_{k,j}(v_{-i}) = 1$ for all $k \neq i$.

(c) Let $f_i = \prod_{k \neq i} u_k \Big/ \prod_{k \neq i} u'_k$.

3. Allocate each item j to each agent i with probability $q_{i,j} = f_i \, p^{\mathrm{NSW}}_{i,j}(v)$.

Note that the fraction f_i of the NSW maximizing assignment allocated to agent i is equal to the relative loss in utility that i's presence imposes on the other agents. The denominator is independent of i's declared valuations, so, in maximizing $f_i \cdot u_i$, which would be agent i's goal, she is maximizing the NSW when she reports truthfully. Cole et al. (2013) show that $f_i \in (1/e, 1]$, without the unit constraint on allocations, but the same argument holds with the unit constraint.

Theorem 16 ((Cole et al., 2013))**.** *The partial allocation mechanism is truthful, feasible, and allocates each agent i at fraction f_i of the NSW maximizing assignment, where f_i is at least* $1/e$.

5.2 The Nash Bargaining Benchmark

In this section, we define our cardinal benchmark as well as an approximation measure for evaluating mechanisms for the one-sided matching problem. Our benchmark is the Nash bargaining solution with a uniformly random matching as the disagreement point. Each agent i's expected utility for this disagreement point is $o_i = \frac{1}{n} \sum_j v_{i,j}$ and the Nash bargaining solution is the outcome p^* that maximizes the Nash Social Welfare objective with respect to the marginal valuations $v - o$. In other words, the Nash bargaining solution distributes the additional value, beyond each agent's outside option, in a fair and efficient manner.

Definition 17. The Nash bargaining solution with disagreement point $(o_i)_{i \in N}$ is

$$p^* = \text{argmax}_p \left\{ \prod_i \Big(\sum_j v_{i,j}\, p_{i,j} - o_i \Big) \right\},$$

where every agent i is constrained to have non-negative utility $\sum_j v_{i,j} p_{i,j} - o_i \geq 0$.

Apart from its fairness properties, this benchmark is also appealing because of its invariance to additive shifts and multiplicative scalings of any agent's values for the items. Shifting all the values of an agent by adding some constant does not affect the marginal values after the outside option is subtracted. Also, scaling all of the values of an agent by some constant does not have any impact on what the Nash bargaining solution, p^*, is; the product value of every outcome is multiplied by the same constant, and hence the optimum is unaffected. As a result, we do not need to assume that the values reported by the agents are scaled in any particular way. One thing to note about the benchmark being invariant to these changes is that, on instances where the agents' values are identical up to shifts and scales, the benchmark assignment is the uniform random assignment.[3]

Our goal is to approximate p^*, the Nash bargaining solution with disagreement point $(o_i)_{i \in N}$, using the following per-agent approximation measure.

[3]The combined property of shift and scale invariance has some counterintuitive implications. Consider an example instance where all agents i have value $v_{i,1} > 1$ for item 1, and $v_{i,j} = 1$ for all other items $j \in \{2, \ldots, n\}$. In the Nash bargaining solution, all agents receive a uniform random item and in particular a $1/n$ fraction of the preferred item 1. This outcome may seem surprising as it does not account for the possibility that some agents may prefer item 1 much more than other agents. This uniform outcome results because the agents' preferences are equivalent up to additive and multiplicative shifts.

Definition 18. The per-agent approximation of mechanism p with respect to benchmark assignment p^* is the worst-case ratio of the utility of any agent in p^* and p. i.e.,

$$\max_v \left\{ \max_i \left\{ \frac{\sum_j v_{i,j} p^*_{i,j}(v)}{\sum_j v_{i,j} p_{i,j}(v)} \right\} \right\}.$$

5.3 Inapproximability by Ordinal Mechanisms

Ordinal mechanisms are popular in the literature on matching. Rather than asking agents for cardinal values for each item, an ordinal mechanism need only solicit an agent's preference order over the items. Two prevalent ordinal mechanisms are the random serial dictatorship (RSD) and probabilistic serial (PS) mechanisms. One of our main motivations for studying cardinal mechanisms in this chapter is that ordinal mechanisms are bound to generate unfair allocations for some instances, due to the fact that they disregard the intensity of the agents' preferences; even when the agents agree, or partially agree, on their preference order, they may still disagree on preference intensities. A mechanism that does not take these intensities into consideration is, for example, unable to distinguish between agents whose favorite item is very strongly preferred over the rest, and agents who have only a slight preference for their top item over the rest.[4] Note that such correlated ordinal preferences may be commonplace in many public sector settings.

Our first lower bound shows that the random serial dictatorship mechanism can be very unfair to some agent, leading to an approximation factor as bad as

[4]For instance, a set of families might each agree on their most-preferred school during a student-to-school matching process, but some families might have a significantly stronger preference for their first choice school due to special programming offered at the school or proximity.

n (the number of agents).

Lemma 19. *The worst case approximation ratio of the random serial dictatorship (RSD) mechanism with respect to the Nash bargaining benchmark is n.*

Proof. Consider the example where agent 1 has value 1 for item 1 and no value for any other item, and each agent $i \geq 2$ has value 1 for item 1, value $1 - \epsilon$ for item i, and no value for other items. That is consider $v =$:

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1-\epsilon & 0 & \dots & 0 \\ 1 & 0 & 1-\epsilon & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 1-\epsilon \end{bmatrix}$$

Table 5.1: Example showing the worst case approximation ratio of the random serial dictatorship mechanism with respect to the Nash bargaining benchmark.

In RSD an ordering of the agents is generated uniformly at random, and then each agent is allocated her favorite available item in that order. In this instance, the first agent in the random ordering will always select item 1, and every agent has the same probability, $1/n$, of being ordered first. Since agent 1 has no value for any other item, the expected utility of this agent 1 in RSD is $1/n$.

On the other hand, as ϵ approaches zero, the Nash bargaining solution assigns each agent i to item i with probability that approaches 1. To verify this fact, note that for $\epsilon = 0$ the Nash bargaining would assign agent i to item i with probability 1, and observe that the distribution that RSD outputs is continuous in ϵ. Thus, the utility of each agent in the Nash bargaining solution – and specifically of agent 1 – approaches 1. As a result, the RSD mechanism is being unfair to agent 1, leading to an approximation factor of n. □

In fact, with a small modification of the instance used to verify how unfair the RSD mechanism can be, the following theorem shows that *every* ordinal mechanism is susceptible to this issue.

Theorem 20. *The worst case approximation ratio of any ordinal mechanism to the Nash bargaining benchmark is at least $n-1$.*

Proof. Consider the following instance v, where agents correspond to rows and items to columns:

$$\begin{bmatrix} 1 & \epsilon & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 1-\epsilon & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & 1-\epsilon & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & \dots & 1-\epsilon & 0 \\ 1 & 0 & 0 & 0 & \dots & 0 & 1-\epsilon \\ 0 & 1 & 1 & 1 & \dots & 1 & 1 \end{bmatrix}$$

Table 5.2: Example showing the worst case approximation ratio of any ordinal mechanism to the Nash bargaining benchmark.

A key property of this instance is that the top $n-1$ agents are ordinally indistinguishable. Each of them ranks item 1 first, one of items $\{2, \dots, n\}$ second, and all other items last. On the other hand, each item $j \in \{2, \dots, n\}$ is ordinally indistinguishable. Each is ranked second by exactly one of the top $n-1$ agents and ranked equivalently by agent n.

Fix an ordinal mechanism. The ordinal indistinguishability of agents $\{1, \dots, n-1\}$ implies, without loss of generality up to agent relabeling, that agent 1 receives item 1 with probability at most $1/(n-1)$. Thus, in the limit of ϵ going to 0, agent 1 obtains a utility of $1/(n-1)$ in this ordinal mechanism.

The Nash bargaining solution is continuous in ϵ and with $\epsilon = 0$ it gives each agent the maximum utility of 1 by allocating item 1 to agent 1, item 2 to agent n, and item $i+1$ to agent i for $i \in \{2, \ldots, n-1\}$. Thus, in the limit, as ϵ goes to zero the Nash bargaining solution gives agent 1 a utility of 1. Therefore, the per-agent approximation of any ordinal mechanism with respect to the Nash bargaining benchmark is $n - 1$. $\square$

5.4 Randomized Partial Improvement

In this section, we define the random partial improvement matching mechanism. This mechanism truthfully elicits the agents' cardinal preferences and uses them in a non-trivial manner to select an outcome. We prove that the per-agent approximation of this mechanism with respect to the Nash bargaining benchmark is proportional to the population monotonicity of the benchmark and its worst-case approximation is significantly better than that of any ordinal mechanism. The approach of the mechanism is to run the PA mechanism with the outside option given by the uniform random assignment on a large sample of the agents and a large fraction of the supply. The resulting mechanism inherits the truthfulness of the PA mechanism.

There are two key difficulties with this approach. First, in order to faithfully implement the outside option, some of the supply needs to be kept aside in the same proportion as the original supply. To enable this set aside, we need to reduce the allocation consumed by the PA mechanism; we achieve this with a novel use of random sampling (Goldberg et al., 2006). Second, it is non-trivial to compare an agent's utility across Nash social welfare maximizing assignments for the original market and a sample of the market. A major endeavor of our analysis shows that per-agent utility is approximately monotone, i.e., the fraction of an agent's utility that is lost as the competition from other agents decreases is non-trivially bounded. (Note, competition from other agents decreases as they are removed from the market.) Our mechanism, then, is structured to take advantage of this approximate monotonicity.

The mechanism is defined by a sequence of steps that gradually construct a doubly-stochastic matrix. By the Birkhoff von Neumann theorem, this matrix corresponds to a probability distribution over matchings. The high-level steps and intuition are as follows: the mechanism samples half the agents and runs at half scale (i.e., with half-unit-demand agents and half-unit-supply items) the PA mechanism with the outside option given by the uniform random assignment. The total demand of half the agents (roughly $n/2$) with half-unit demand is a quarter of the total supply (roughly $n/4$), so there is a leftover $n/4$ of supply from the half-units on which the PA mechanism was run. A further one-quarter of each of the n units is used to provide a half-unit of the outside option to each of the (roughly) $n/2$ agents in the sample. The final quarter is used to replace as necessary the fractions of items withheld due to the fractional reduction in the PA mechanism. Necessarily, the one-unit allocation to these agents uses up half the supply. The remaining half of the supply is then allocated recursively to the remaining half of the agents.

A formal description of this mechanism is given below. Since we call this mechanism recursively after some agents' allocation has been finalized (in the form of marginal probabilities) and some portions of the items have been allocated, we define it for the remaining $\bar{n} \leq n$ agents and the original n items whose capacities may have been reduced from 1 to $(c_i)_{i \in M} \in [0, 1]^n$.

Definition 21. Given some value $n_0 \in \mathbb{N}$, the *randomized partial improvement*

(RPI) mechanism on $\bar{n} \leq n$ agents and n items with supplies $c_1, \ldots, c_n$ such that $\sum_{j=1}^{n} c_j = \bar{n}$ works as follows:

1. If $\bar{n} < n_0$, allocate the remaining item capacities uniformly at random, i.e., return $p_{i,j} = c_j/\bar{n}$ for each agent i and terminate. Otherwise, continue.
2. Randomly sample a subset N' of $n' = \lceil \bar{n}/2 \rceil$ agents.
3. On the sampled agents run the PA mechanism with the outside option given by the uniform random assignment $o'_i = \frac{1}{\bar{n}} \sum_{j=1}^{n} v_{i,j} c_j$ from the supplies. Denote the allocation of item j to agent i by $q'_{i,j}$; the total amount allocated to agent i is $f'_i = \sum_{j=1}^{n} q'_{i,j}$.
4. Allocate to each $i \in N'$ half of their PA assignment and "pad" it with the outside option to ensure a unit allocation. As a result, the total allocation of item j to agent i is $p'_{i,j} = q'_{i,j}/2 + (1 - f'_i/2)\, c_j/\bar{n}$.
5. Recursively run RPI on the remaining $n'' = \bar{n} - n'$ agents and item supplies $c''_j = c_j - \sum_{i \in N'} p'_{i,j}$.
6. Return the assignment p that combines the assignment p' for N' and the assignment p'' returned by the recursive call for the remaining n'' agents.

The following proof of correctness (feasibility and truthfulness) formalizes the intuition preceding the definition of the mechanism. The ideas of the proof are more transparent in the case where $\bar{n}$ is even and, in particular, $\lceil \bar{n}/2 \rceil = \lfloor \bar{n}/2 \rfloor = \bar{n}/2$.

Theorem 22. *The randomized partial improvement mechanism with $n_0 \geq 4$ on n unit-demand agents and n unit-supply items is feasible. i.e., it gives fractional allocations that produce a doubly stochastic matrix, and truthful. i.e., it is a dominant strategy for each agent to truthfully report her value for each item.*

Proof. Feasibility is shown by induction on the recursive definition of the mechanism. The inductive hypothesis is that the fractional allocation on $\bar{n}$ agents with supplies $c_1, \ldots, c_n$ such that $\sum_j c_j = \bar{n}$ has a total fractional allocation to each agent of one, i.e., $\sum_j p_{i,j} = 1$ for each i, and a total fractional allocation of each item equal to its supply, i.e., $\sum_i p_{i,j} = c_j$ for each j. The base case of $\bar{n} < n_0$ clearly satisfies the inductive hypothesis. For the inductive step, the key point to argue is that the supply c_j of each item j is sufficient to cover the allocation to the sampled agents.

This can be seen as follows: the $\lceil \bar{n}/2 \rceil$ sampled agents are allocated half of their PA assignment on the supplied capacities. Since $f'_i \geq \frac{1}{e}$ for all i by Theorem 16, this means that the amount of each agent's half PA assignment is $\sum_j q'_{i,j}/2 = f'_i/2 \geq 1/(2e)$. To ensure that each agent gets exactly one item in expectation, Step 4 pads this allocation with a uniformly random assignment, which may thus require up to $1 - 1/(2e)$ units for each of the $\lceil \bar{n}/2 \rceil$ sampled agents, for a total of $(1 - 1/(2e))\lceil \bar{n}/2 \rceil$. But, since the full PA assignment allocated no more than c_j of each item j, the half PA assignment set aside at least $c_j/2$ of each item, leading to a total of $\sum_j c_j/2 = \bar{n}/2$. We conclude by observing that $(1 - 1/(2e))\lceil \bar{n}/2 \rceil$ is at most $\bar{n}/2$ when $\bar{n} \geq n_0 = 4$, so the amount set aside from each item is sufficient to cover the sampled agents' allocation.

Truthfulness follows by considering each agent conditioned on the state of the mechanism during the recursive step where that agent is selected in the sampled set N'. The agent's report plays no role in determining the state at this point. Given the state, the outcome for this agent is fully determined by the PA mechanism which is truthful. Thus, the mechanism is truthful. □

To bound the per-agent utility of the RPI mechanism, we analyze the contribution to the utility of an agent who is sampled in the outermost recursive call of the mechanism. An agent is sampled as such with probability at least one half, and otherwise the agent's utility is at least zero. The utility of these sampled agents is easily compared to the utility of the PA mechanism (without the agents that are not sampled). An issue significantly complicating the analysis of the approximation is the fact that we need to compare the utility of an agent sampled in this invocation of the PA mechanism with their utility in the Nash bargaining solution on the full set of agents. Counter-intuitively, it is not true that these agents are always better off without the competition from the agents that are not sampled: there are instances where removing some of the competition, in fact, lowers the utility of an agent.

In Section 5.5, we define the *ρ-utility monotonicity for NSW* to be the maximum non-monotonicity of utility of any agent and sets of agents N and subset N' with NSW maximizing solutions p and p' respectively. i.e.,

$$\rho := \max_{N' \subseteq N} \left\{ \max_{i \in N'} \left\{ \frac{\sum_j v_{i,j} p_{i,j}}{\sum_j v_{i,j} p'_{i,j}} \right\} \right\}.$$

This parameter quantifies the extent to which some agent may be worse off in the NSW maximizing solution after the removal of some subset of agents. Defining the worst-case value of ρ across instances and subsets as ρ^*, Section 5.5 bounds ρ^* between $\Omega(2^{\sqrt{\log n}/2})$ and $O(2^{2\sqrt{\log n}})$, the latter of which is $o(n^\epsilon)$ for any constant $\epsilon > 0$. It is worth noting that we ran experiments on a large set of instances including a range of natural utility functions and found that this value was actually no more than 1 in all of these instances.

Theorem 23. *The randomized partial improvement mechanism with $n_0 = 4$ on n unit-demand agents and n unit-supply items is a $4\,e\,\rho$-approximation to the Nash bargaining solution with disagreement point given by the uniform random assignment.*

Proof. If $n < n_0 = 4$ then the base case of RPI is invoked and a uniform assignment is returned. With $n < 4$, however, this assignment is a $3 < 4\,e\,\rho$-approximation, as each agent obtains $1/3$ of each item.

Otherwise, we analyze the contribution to the utility of an agent conditioned on the agent being sampled in the first recursive call of the algorithm. This event happens with probability at least $1/2$. When this happens the utility of the agent is half the utility of PA on the sampled agents plus half the utility from the outside option. The ρ-utility monotonicity property implies that the utility of an agent in the NSW maximizing outcome on the sample is a ρ approximation to the same agent's utility in the NSW maximizing outcome on the full set of agents. Running PA guarantees an e fraction of this utility. Combining these steps and the fact that agents who are not sampled in the first recursive call still receive nonnegative utility, we obtain a $4\,e\,\rho$-approximation. □

Combined with Theorem 28 from Section 5.5 which bounds ρ^* by $O(2^{2\sqrt{\log n}})$, we have the following result:

Corollary 24. The randomized partial improvement mechanism with $n_0 = 4$ on n unit-demand agents and n unit-supply items guarantees an approximation of the Nash bargaining solution with uniform outside option with a factor $O(2^{2\sqrt{\log n}}) \subseteq o(n^{\epsilon})$ for any constant $\epsilon > 0$.

5.5 Approximate Utility Monotonicity

A factor significantly complicating the analysis of the approximation of the random partial improvement mechanism is the fact that the benchmark is computed based on the Nash social welfare maximizing solution when all agents in N are present, while the mechanism's performance depends on the solution for N', the sampled agents. The NSW maximizing solution for N and N' can generally be quite different. Moreover, as it turns out, there are instances where the utilities of some agents in the NSW maximizing solution are non-monotone with respect to removal of other agents, i.e., there exist instances that exhibit positive externalities between agents. Table 5.3 gives a simple example of such an instance (discussed in detail later on) and the remainder of the section develops upper and lower bounds on the worst-case non-monotonicity of utility.

	A	B	C
a	1	2	0
b	0	2	1
c	0	0	1

i. Agent valuations

A	B	C
1	0	0
0	1	0
0	0	1

ii. Initial solution

A	B	C
1/2	1/2	0
0	1/2	1/2

iii. Final solution

Table 5.3: A simple instance involving three agents; a, b, c; and three items; A, B, C. The value of agent b in the Nash bargaining solution after the removal of agent c drops by a factor of 4/3. The optimality of solutions (ii) and (iii) for Nash social welfare are intuitive; however, formal justification is given in the lower bound section where this example is revisited.

Definition 25. A matching environment on agents N is *ρ-utility monotone* if for any subset N' of N and any $i \in N'$ the utility of i in the NSW maximizing assignment, p', for N' is at least a ρ-approximation to the NSW maximizing assignment, p, for N:

$$\rho := \max_{N' \subseteq N} \left\{ \max_{i \in N'} \left\{ \frac{\sum_j v_{i,j} p_{i,j}}{\sum_j v_{i,j} p'_{i,j}} \right\} \right\}.$$

This parameter quantifies the extent to which some agent may be worse off in the NSW solution after the removal of some subset of agents. We let ρ^*, denote the worst case value of ρ across instances; this value is known as the *opportunity structure* of the Nash bargaining solution for this class of instances (Thomson and Lensberg, 1989). In Section 5.5.1 we prove an upper bound of $O(2^{2\sqrt{\log n}})$, which is $o(n^{\epsilon})$ for any constant $\epsilon > 0$, for the value of ρ^* over all one-sided matching instances, and in Section 5.5.2 we complement this result by proving a lower bound of $\Omega(2^{\sqrt{\log n}/2})$ for this value.

5.5.1 Upper Bound

Given a valuation matrix v and a random matching p, we henceforth use $u_i(p)$ to denote the expected utility of agent i for p given v, i.e., $\sum_{j \in M} v_{i,j} p_{i,j}$ (similarly, we use $u_i'(p)$ for valuation matrix v'). Note that, as we discussed in Section 5.2, the Nash bargaining solution is scale invariant. Therefore, if we scale the valuations of each agent i by some constant $c_i > 0$, then the Nash bargaining solution with respect to valuations $c_i v_{i,j}$ instead of $v_{i,j}$ will remain the same. This means that given some problem instance that yields a doubly stochastic matrix p as its Nash bargaining solution, we can always "normalize" the valuations of the agents so that every agent's expected utility for p is equal to 1, and p remains that Nash bargaining solution of the normalized instance. This is a convenient normalization that we make use of below.

In order to prove the upper bound on ρ^*, we first prove the following useful lemmata.

Lemma 26. *Let p be a NSW maximizing solution, and v be the valuations normalized so that for every agent i, $u_i(p) = 1$. Then, if some agent i is allocated an item j with positive probability, i.e., $p_{i,j} > 0$, every other agent $k \neq i$ must have $v_{k,j} \leq v_{i,j} + 1$. Equivalently, $v_{i,j} \geq \max_{k \in N}\{v_{k,j}\} - 1$.*

Proof. For contradiction, assume that there exist two agents k and i and an item j such that $p_{i,j} > 0$ and $v_{k,j} = v_{i,j} + 1 + \delta$ for some $\delta > 0$. Since the expected utility of agent k is 1, there must also exist some item ℓ with $p_{k,\ell} > 0$ and $v_{k,\ell} \leq 1$ (otherwise the expected utility of agent k would be greater than 1). Note that $\ell \neq j$, since $v_{k,j} = v_{i,j} + 1 + \delta > 1$, whereas $v_{k,\ell} \leq 1$.

Let p' be a probability distribution identical to p, except $p'_{k,j} = p_{k,j} + \epsilon$, $p'_{i,j} = p_{i,j} - \epsilon$, $p'_{k,\ell} = p_{k,\ell} - \epsilon$, and $p'_{i,\ell} = p_{i,\ell} + \epsilon$, for some positive $\epsilon < \min\{p_{i,j}, p_{k,\ell}\}$, whose exact value we choose later on. In other words, p' swaps probability ϵ between agents i, k and items j, ℓ. The expected utility of agent k in p' is

$$\begin{aligned} u_k(p') &= 1 + \epsilon v_{k,j} - \epsilon v_{k,\ell} \\ &\geq 1 + \epsilon(v_{i,j} + 1 + \delta) - \epsilon \\ &= 1 + \epsilon v_{i,j} + \epsilon\delta, \end{aligned}$$

and the expected utility of agent i in p' is

$$\begin{aligned} u_i(p') &= 1 + \epsilon v_{i,\ell} - \epsilon v_{i,j} \\ &\geq 1 - \epsilon v_{i,j}. \end{aligned}$$

Since every other agent's expected utility is the same in p' and p (equal to 1), the NSW of p' is

$$\begin{aligned} \prod_{i \in N} u_i(p') &\geq (1 + \epsilon v_{i,j} + \epsilon\delta)(1 - \epsilon v_{i,j}) \\ &= 1 + \epsilon(\delta - \epsilon(v_{i,j}^2 + \delta v_{i,j})). \end{aligned}$$

Therefore, if we let $\epsilon < \delta/(v_{i,j}^2 + \delta v_{i,j})$), the NSW of p' is greater than 1, which is the NSW of p, contradicting the fact that p is a NSW maximizing solution. □

Lemma 27. *Given a problem instance, let p and $\overline{p}$ be the NSW maximizing outcomes before and after (respectively) some subset of the agents has been removed. If among the* *remaining agents there exists a set of agents N_1 and a constant $d \geq 12$ such that every agent $i \in N_1$ has $u_i(\overline{p}) \leq u_i(p)/d$, then there also exists a larger set N_2 of remaining agents such that $|N_2| \geq d|N_1|/3$ and for all agents $k \in N_2$ we have $u_k(\overline{p}) \leq 4u_k(p)/d$.*

Proof. Without loss of generality, let v be the agent valuations normalized so that $u_i(p) = 1$ for every agent i, and v' be the valuations normalized so that $u'_i(\overline{p}) = 1$. Given the $v_{i,j}$ values that yield $u_i(p) = 1$, we can get the $v'_{i,j}$ values that yield $u'_i(\overline{p}) = 1$ using the simple formula $v'_{i,j} = v_{i,j} \cdot \frac{u_i(p)}{u_i(\overline{p})}$. In other words, for each agent i who is worse-off in $\overline{p}$ compared to p, i.e., $u_i(\overline{p}) < u_i(p)$, we scale all of that agent's item values up by the same factor, $u_i(p)/u_i(\overline{p})$. In particular, for each agent $i \in N_1$ this means that $v'_{i,j} \geq dv_{i,j}$ for every item j.

For every $i \in N_1$ we know that the drop in that agent's value with respect to the original valuations is $u_i(p) - u_i(\overline{p}) \geq u_i(p)(1 - 1/d) = 1 - 1/d$. In order to account for that drop, we partition the set of items of which i is allocated more in p compared to $\overline{p}$ into two sets depending on whether $v_{i,j} \geq 0.5$ or not: $M_i^h = \{j \in M : p_{i,j} > \overline{p}_{i,j} \text{ and } v_{i,j} \geq 0.5\}$ and $M_i^\ell = \{j \in M : p_{i,j} > \overline{p}_{i,j} \text{ and } v_{i,j} < 0.5\}$. We first show that from the aforementioned $1 - 1/d$ drop in value, no more than 0.5 could be due to the items in M_i^ℓ, since

$$\sum_{j \in M_i^\ell} (p_{i,j} - \overline{p}_{i,j}) v_{i,j} \; < \; 0.5 \sum_{j \in M_i^\ell} (p_{i,j} - \overline{p}_{i,j}) \; \leq \; 0.5.$$

Therefore, at least $0.5 - 1/d$ of this drop in value for each agent $i \in N_1$ is due to items in M_i^h. Summing this up over all the agents in N_1, we get

$$\sum_{i \in N_1} \sum_{j \in M_i^h} (p_{i,j} - \overline{p}_{i,j}) v_{i,j} \; \geq \; \sum_{i \in N_1} \left(0.5 - \frac{1}{d}\right) \tag{5.1}$$

$$= \; |N_1| \left(0.5 - \frac{1}{d}\right). \tag{5.2}$$

Let $N_2 = \{k \in N : \overline{p}_{k,j} > 0 \text{ for some } j \in \bigcup_{i \in N_1} M_i^h\}$ be the set of agents that are allocated with positive probability in $\overline{p}$ an item from M_i^h for some $i \in N_1$. Using Lemma 26 we get that for every item $j \in M_i^h$, if $\overline{p}_{k,j} > 0$ then $v_{k,j} \leq v_{i,j} + 1$ and $v'_{k,j} \geq v'_{i,j} - 1$. Using the fact that $v'_{i,j} \geq dv_{i,j}$ for every $i \in N_1$, shown above, the latter inequality also implies that $v'_{k,j} \geq dv_{i,j} - 1$. Therefore

$$\frac{v'_{k,j}}{v_{k,j}} \geq \frac{dv_{i,j} - 1}{v_{i,j} + 1} \geq d - \frac{d+1}{1.5} \geq \frac{d}{4},$$

where the second inequality uses the fact that $v_{i,j} \geq 0.5$ and the last inequality uses the fact that $d \geq 8$. This implies that for every $k \in N_2$

$$\begin{aligned} u_k(\overline{p}) &= \sum_{j \in M} \overline{p}_{k,j} v_{k,j} \\ &\leq \sum_{j \in M} \overline{p}_{k,j} \frac{4}{d} v'_{k,j} \\ &= \frac{4}{d} \sum_{j \in M} \overline{p}_{k,j} v'_{k,j} \\ &= \frac{4u'_k(p)}{d} \\ &= \frac{4u_k(p)}{d}, \end{aligned}$$

where the last equation uses the fact that $u'_k(p) = u_k(p) = 1$ according to our normalization.

Since we have shown that for all $k \in N_2$ we have $u_k(\overline{p}) \leq 4u_k(p)/d$, it now suffices to show that the size of N_2 is at least $d|N_1|/3$. Since, for any item $j \in \bigcup_{i \in N_1} M_i^h$, any agent k with $\overline{p}_{i,j} > 0$ satisfies $v'_{k,j} \geq dv_{i,j} - 1$, the total value, with respect to valuations v', generated by the item fractions of the items "lost" by the agents in N_1 is at least

$$
\begin{aligned}
\sum_{i \in N_1} \sum_{j \in M_i^h} & (p_{i,j} - \overline{p}_{i,j})(dv_{i,j} - 1) \\
& \geq d|N_1|\left(0.5 - \frac{1}{d}\right) - \sum_{i \in N_1} \sum_{j \in M_i^h} (p_{i,j} - \overline{p}_{i,j}) \\
& \geq d|N_1|\left(0.5 - \frac{1}{d}\right) - |N_1| \\
& \geq \frac{d-4}{2}|N_1| \\
& \geq \frac{d}{3}|N_1|,
\end{aligned}
$$

where the last inequality uses the fact that $d \geq 12$. But, since the total value of each agent in N_2 with respect to valuations v' is exactly 1, there need to be at least $\frac{d}{3}|N_1|$ agents in N_2 sharing this value, otherwise there would exist some agent $i \in N_2$ such that $u'_i(\overline{p}) > 1$. □

Theorem 28. *For any problem instance, the value of ρ is $O(2^{2\sqrt{\log n}}) \subseteq o(n^\epsilon)$ for any constant $\epsilon > 0$.*

Proof. In order to prove this bound, we will repeatedly apply the result of Lemma 27. Let p and $\overline{p}$ be the NSW maximizing outcomes in a problem instance before and after some subset of the agents has been removed and, without loss of generality, let v be the agent valuations normalized so that $u_i(p) = 1$ for every agent i, and v' be the valuations normalized so that $u'_i(\overline{p}) = 1$.

By Definition 25, in an instance with utility monotonicity equal to ρ, there exists at least one agent $i \in N_1$ such that $u_i(p)/u_i(\overline{p}) = \rho$ or $u_i(\overline{p}) = u_i(p)/\rho$. If $\rho > 12$, then Lemma 27 would imply that there also exists a set N_2 of at least $\rho/3$ agents such that $u_k(\overline{p}) \leq 4u_k(p)/\rho = 4/\rho$ for every $k \in N_2$. Lemma 27, combined with the existence of the set N_2, in turn, implies the existence of an even larger group

N_3 of at least $(\frac{1}{3}\rho) \cdot (\frac{1}{3}\rho/4)$ agents, and each agent $k \in N_3$ has value $u_k(\overline{p}) \leq 16/\rho$. Applying Lemma 27 a total of α times thus implies the existence of a set of at least $(\rho/3)^\alpha \cdot (1/4)^{\alpha(\alpha-1)/2}$ agents such that each such agent k has value $u_k(\overline{p}) \leq 4^\alpha/\rho$. Assume that there exists some instance for which ρ is at least $4^{\sqrt{\log n}+1}$. If we choose $\alpha = \sqrt{\log n}$, however, this implies the existence of $(\rho/3)^\alpha \cdot (1/4^{\alpha(\alpha-1)/2}) \geq 4^{(\sqrt{\log n}+1)\cdot\sqrt{\log n}}/[(3/2)^{\sqrt{\log n}} \cdot 2^{\sqrt{\log n}\cdot\sqrt{\log n}}] \geq (8/3)^{\sqrt{\log n}} \cdot n \geq n$ agents of value at most $4^\alpha/\rho \leq 1/4$. But, this would imply that all the agents have a value less than 1 in $\overline{p}$, which contradicts the fact that $\overline{p}$ is a NSW maximizing solution because the product in p is equal to 1. □

5.5.2 Lower Bound

We conclude with a lower bound showing that for a very carefully designed family of instances, the upper bound of Theorem 28 is essentially tight.

Theorem 29. *There exists a family of problem instances for which $\rho^* = \Omega(2^{\sqrt{\log n}/2})$.*

Due to the complexity of the construction that yields Theorem 29, we defer its description to Appendix B.1. To exhibit how we use KKT conditions to prove that this elaborate construction implies the desired bound, we use the rest of this section to apply this approach to the much simpler construction of the example in Table 5.3, which yields a bound of $\rho^* \geq 4/3$.

Our lower bound construction in the appendix proceeds by building a family of instances (parameterized by the number of agents n), and in each instance, we define an "initial" setting in which all agents are present, and a "final" setting, in which some agents have been removed. For each setting, we identify the

Nash bargaining solution, respectively called the initial and final solution. We focus on a particular agent, called the *loser*, who is present in both settings. We show that the loser's valuation drops by a multiplicative factor μ in going from the initial to the final solution, and consequently, $\rho \geq 1/\mu$ for that market and $\rho^* \geq 1/\mu$ overall.

To prove a lower bound, we need to be able to verify that a given doubly stochastic matrix is indeed the Nash bargaining solution of the instance at hand. We do so using the KKT conditions, which allow us to interpret these solutions as a form of market equilibrium. The optimization problem which yields the Nash bargaining solution in one-sided matching markets, is shown below (where $m = n$ is used to denote the number of items):

$$\max \sum_{i=1}^{n} \log \left[\sum_{j=1}^{n} v_{i,j} p_{i,j} \right]$$

$$\text{such that:} \quad \text{for all } i: \quad \sum_{j=1}^{m} p_{i,j} \leq 1$$

$$\text{for all } j: \quad \sum_{i=1}^{n} p_{i,j} \leq 1$$

$$\text{for all } i, j: \quad p_{i,j} \geq 0.$$

If t_j is the dual variable related to each item j, and q_i is the dual variable related to each agent i in the above program, then the KKT conditions state that

$$\text{for all } j: \; t_j \geq 0, \; \text{ and } \; t_j > 0 \Longrightarrow \sum_{i=1}^{n} p_{i,j} = 1 \tag{5.3}$$

$$\text{for all } i: \; q_i \geq 0 \; \text{ and } \; q_i > 0 \Longrightarrow \sum_{j=1}^{m} p_{i,j} = 1 \tag{5.4}$$

$$\begin{gathered} \text{for all } i, j: \quad \frac{v_{i,j}}{t_j + q_i} \leq \sum_{j=1}^{m} v_{i,j} p_{i,j} \; \text{ and} \\ p_{i,j} > 0 \Longrightarrow \frac{v_{i,j}}{t_j + q_i} = \sum_{j=1}^{m} v_{i,j} p_{i,j} \end{gathered} \tag{5.5}$$

	A	B	C
a	1	2	0
b	0	2	1
c	0	0	1

i. Agent valuations

	A	B	C	q
	1	2	0	1
	0	**1**	1/2	0
	0	0	**1**	0
t	0	1	1	

ii. Initial solution

	A	B	C	q
	2/3	**4/3**	0	0
	0	**4/3**	**2/3**	0
t	2/3	4/3	2/3	

iii. Final solution

Table 5.4: Simple instance involving three agents: a, b, c; and three items; A, B, C. The value of agent b in the Nash bargaining solution after the removal of agent c drops by a factor of $4/3$. Normalized values are depicted in (ii) and (iii)) along with prices t and q; these values are depicted in bold-face if the allocation probability of the NSW solution is non-zero.

The KKT conditions are necessary and sufficient for the optimal solution when the constraints are linear and the objective is convex, as is the case here. To check whether a given candidate solution p is a Nash bargaining solution for some instance, we first normalize the valuations so that $\sum_{j=1}^{m} v_{i,j} p_{i,j} = 1$ for all i. Then, at a solution satisfying the KKT conditions we have $v_{i,j} = t_j + q_i$ if $p_{i,j} > 0$ and $v_{i,j} \leq t_j + q_i$ if $p_{i,j} = 0$. Thus, a solution that satisfies these two conditions plus conditions (5.3)–(5.4) is a Nash bargaining solution. Based on these conditions, the values of t_j can be interpreted as item-specific "prices" and the values of q_i as agent-specific "prices", leading to an interpretation of the Nash bargaining solution as a market equilibrium: to "buy" a $p_{i,j}$ fraction of item j, agent i needs to spend $(t_j + q_i)p_{i,j}$, and each agent prefers to buy only items with the best value over price ratio (see condition (5.5)).

To illustrate the usefulness of these variables, which are used extensively in the appendix, we revisit the instance of Table 5.3 where the items are named A, B, and C; the bidders a, b, and c; and the unscaled valuations of the agents appear in Table 5.4 (i).

First, we observe that in the initial equilibrium (with all agents present), a, b, c receive items A, B, C, respectively, each with probability 1. In Table 5.4(ii) we show the normalized values of the agents in this equilibrium and we also provide the dual variables t_j for each item j and q_i for each agent i. It is easy to verify that the aforementioned KKT conditions are satisfied in this case and hence this is indeed the Nash bargaining solution when all agents are present. If agent c is removed, then the final equilibrium finds a receiving each of A and B with probability $\frac{1}{2}$, while b receiving each of B and C with probability $\frac{1}{2}$. Table 5.4(iii) provides the scaled valuations and dual variable values for this outcome, and it is again easy to verify that KKT conditions are satisfied. In this example, bidder b is the loser. Using the valuations from Table 5.4(ii), her value in the initial equilibrium was 1 and it dropped to 0.75 in the final equilibrium, leading to $\rho = \frac{4}{3}$ in this example.

5.6 Discussion

In this chapter, we defined the random partial improvement (RPI) mechanism for one-sided matching markets without monetary transfers. RPI truthfully elicits the cardinal preferences of the agents, assigns each agent to an item, and outputs a distribution over matchings that approximates every agent's utility in the Nash bargaining solution. Mechanisms assignment problems, such as the house allocation problem, are used in a variety of contexts, including in public sector settings where individuals and families are to be assigned to seats in classrooms, low-income housing assistance programs, as well as assignment of human resources such as doctors and teachers. There maybe hard constraints in this setting – such as that we would like to assign each individual to a house or a doctor to a hospital – but we would also like to do so satisfying certain truthfulness, fairness, and efficiency criteria. The mechanism presented here adds to this body of work by leveraging a random sampling technique.

We present a set of open questions and discussions related to our mechanism and the house allocation problem below and provide a broader discussion and a series of mechanism design questions inspired by public sector settings in Part V. Our analysis of the mechanism proposed in this chapter suggests several open questions and directions for future work. A natural open question is whether there exists a truthful mechanism that can achieve a constant factor approximation of the Nash bargaining benchmark. The main obstacle for the RPI mechanism was the non-monotonicity of the Nash bargaining benchmark, so it would be interesting to see if some other mechanism could circumvent this issue. Alternatively, since the construction leading to the lower bound is quite artificial, are there any natural assumptions regarding the valuations of the agents that would mitigate the non-monotonicity?

Another interesting direction would be to study how the utilities of agents in the CEEI outcomes compare to those of the Nash bargaining solution. Recall that the CEEI and the Nash bargaining solution are equivalent in linear markets without the matching constraint (Vazirani, 2007), but are different for matching markets.

Our work provides a non-trivial mechanism aiming to approximate a well-motivated ex-ante Pareto efficient outcome. One could also consider the design of truthful mechanisms aiming to approximate alternative benchmarks on the ex-ante Pareto frontier. Natural candidates would be the utilitarian (or the egalitarian) outcome which maximize the sum (or the minimum) of the agents' utilities. One drawback of these outcomes is that, unlike the Nash bargaining solution, they are not scale invariant, but one could consider scaled variants of their objectives, e.g., where the agent values are normalized so that $\sum_{j \in M} v_{i,j} = 1$ for every agent i.

And, finally, we observe in simulations that RPI outperforms or gives comparable performance to RSD and PS, which are two widely-used and studies ordinal mechanisms in a broad range of settings. Specifically, it does so when the ordinal preferences of agents are correlated but the cardinal intensity of these preferences vary. These simulations are conducted on synthetic data, and it remains an open question how well our mechanism does in data generated by real instances. We can also test, by simulation how well our mechanism minimizes envy-freeness and other fairness properties compared to these mechanisms. A broad set of questions related to matching mechanisms is providing a cross-demographic comparison of the performance: i.e., are there specific characteristics to reports from certain under-served and disadvantaged communities, and, if so, do truthful mechanisms tend to narrow disparities in performance? How about ordinal versus cardinal mechanisms? How about mechanisms that are simpler to explain?

Part III

Improving Access to Information

CHAPTER 6

OVERVIEW OF PART III

In Part III, we discuss how we can develop and deploy computational and network-based techniques to improve access to information. This setting is distinct from that in Part II since we assume that information is not a finite or scarce resource. Nonetheless, identifying what information individuals and communities need is made challenging by numerous other phenomena. We focus on two salient issues here:

In Chapter 7, we note that due to data inequalities, there is a lack of knowledge on the information needs of disadvantaged communities. In data sets that can shed light on health information needs, communities on the margins of society are often under-represented, mis-represented, or entirely missing. Focusing on health in Africa, we explore how to use search data to help bridge this information gap by studying what information individuals based in the continent seek related to HIV/AIDS, malaria, and tuberculosis. We use a mix of computational, topic modeling, and information retrieval techniques to bring to the surface the varied and unmet information needs of individuals. We further show that there is are discrepancies in the quality of content provided to users on search engines depending on the topic with which their searches are associated. We provide an explanation of the policy and societal implications of these findings. To our knowledge, this is the first-ever research study to use large web and social media-based data to study health across all 54 nations in Africa.

In Chapters 8 and 9, we focus on two challenges that remain even after identifying the information needs of communities and individuals. In Chapter 8, we study the *overexposure* phenomena, where encountering information to which one is not receptive may result in a negative outcome. For instance, in anti-smoking campaigns, studies have shown that individuals exhibit varying levels of susceptibility to different types of messaging and that exposure to ill-suited campaign can in a sense "back-fire" (Wakefield et al., 2003). In light of these findings, we revisit the classical *influence maximization* problem in a setting where the objective function incorporates these negative outcomes of overexposure. We present a polynomial-time algorithm for targeting seed agents on a network when we are not constrained by the number of individuals that can belong in the seed set. On the other hand, we show that in the budgeted setting, it is NP-hard to decide if there is a budgeted set yielding a positive payoff.

The *diffusion of innovation* literature has leveraged the structure of networks to spread information that can improve societal welfare across communities. Such efforts can be hampered by the structure of networks. In particular, in segregated networks, information that easily spreads in one part of a network may not reach another part. Often, this may be to the detriment of disadvantaged groups that may not have access to the "core" part of the network. Segregation is facilitated by homophily, or the tendency for individuals to form ties with others that are similar to them. To push back against homophily, we seek to harness the power of organic forces that may already exist and may be pushing against segregation. In Chapter 9, we show that *triadic closure*, which has long been believed to amplify segregation may, in fact, be a force that can alleviate it. We analyze several fundamental network formation models to show supporting evidence and present interventions that can nudge networks towards a more integrated state informed by these findings. We discuss open questions related to networks, inequality, and access to information in the corresponding discussion sections and Part V.

CHAPTER 7

MEASURING ACCESS TO HEALTH INFORMATION IN DATA-SPARSE REGIONS

New technologies and data-sources are constantly being leveraged to upgrade and supplement the design, monitoring, and evaluation of health policy in the developed world. There is, however, a substantial gap in the availability and quality of health data between developing and developed nations. In many developing nations, even when health-related information is collected, it is often neither comprehensive nor digitized. A 2014 regional report by the African Union highlights this issue, noting:

> Unless gaps are identified early and accurately, simply providing a raft of general interventions will not meet the real health needs of the people in the Region (Sambo, 2014).

This *data inequality* can be a roadblock to identifying major public health concerns and implementing effective interventions. While targeted education addressing individuals' health needs is a critical tool for combating disease, health organizations and policy makers struggle to identify what knowledge individuals in developing nations seek and whether their health information needs are being met. It is especially urgent to understand how such needs vary by region and demographic groups since the impact of diseases – their prevalence, progression, and transmission rates – as well as people's disease knowledge and attitudes vary regionally and demographically. Limited understanding of the health information needs of individuals hinders the efficacy of targeted education efforts such as gender- and age-specific programming (De Bruyn, 2000; Germain, 2009; Global Fund, 2016; UNDP, 2015a,b).

In this chapter, we take a step towards narrowing this gap, focusing on the problem of identifying and measuring people's everyday health information needs, concerns, and misconceptions in the African continent. We use Bing search queries originating in all 54 African nations to explore which themes related to infectious disease people are most interested in getting information about, as evidenced by their searches. We focus on HIV/AIDS, malaria, and tuberculosis because, together, these three diseases account for 22% of the disease burden in sub-Saharan Africa (IHME, 2016).

Search data provide a wealth of information on people's real-time activities, experiences, concerns, and misconceptions relatively cheaply (Kern et al., 2016; Paul and Dredze, 2017), allowing us to obtain potentially hard-to-survey information in a bottom-up manner. In contrast, most data-driven efforts aimed at mitigating the impact of disease in data-sparse regions, including the Global Burden of Disease Study and the African Health Observatory, have used a top-down approach, actively collecting data with a particular goal in mind (AHO, 2010; IHME, 2016). Such approaches, while helpful, are often limited in their ability to provide a thorough and comprehensive overview of people's information needs, attitudes, and misconceptions. Existing bottom-up solutions to this problem, such as the West Africa Health Organization's study of health information needs in West Africa, primarily make use of manual interviews (Allen et al., 2010). These approaches can obtain a comprehensive picture of individuals' needs, but are difficult to scale, expensive, and time-consuming. Analyzing search data is a natural candidate for scaling up and complementing studies not only because it addresses some of these challenges, but also because search logs have already been shown to contain large quantities of information related to serious conditions in other contexts (De Choudhury et al., 2014; Paul and Dredze, 2017). Despite the fact that Internet penetration in Africa is growing rapidly—31% of the population is currently covered, with nearly 8,500% growth since 2000 (ITU, 2017)—to our knowledge, no prior work has looked specifically at search data to understand health information needs in all African nations.

Our Results

We analyze Bing search data related to HIV/AIDS, malaria, and tuberculosis from all 54 African nations. We uncover themes in which individuals are interested using latent Dirichlet allocation (LDA), a standard generative model for automatically extracting topics from text (Blei et al., 2003).[1] The topics that emerge cover basics such as symptoms, testing, and treatment, as well as hard-to-survey topics such as stigma and discrimination, beliefs in natural cures and remedies, and concerns about the impact of gender inequality in HIV transmission. We explore the ways in which the popularity of these topics vary by age, gender, and location. We expose patterns including that searches related to pregnancy and breastfeeding and relatively more popular among women while searches related to cure news are relatively more popular among men.

Delving into the content returned to users, we compare the organic search results returned for different topics and quantify the discrepancies in the quality of information returned to users. These results highlight unmet health information needs, concentrated misinformation related to specific health topics, and differences in user satisfaction by topic. We discuss the limitations of our approach, including the difficulty of extrapolating our observations to the wider

[1] We include discussions on measures taken to quantify and mitigate bias in this data collection as well as limitations in the discussion section as well as the corresponding appendix.

population of Africa, and the danger of overlooking the health concerns of communities who are not on the web. Finally, we highlight potential implications of these analyses on health policy and education efforts both on- and off-line.

7.1 Data and Methodology

To generate the data set of HIV/AIDS queries, we first obtained all Bing search queries containing at least one of the terms “HIV” or “AIDS” that originated in any of the 54 African nations between January 2016 and June 2017. We consider this time-period since data at the level of granularity described here was available for this time period. Both mobile and desktop searches were retrieved. Each query record in the data consisted of the raw search query, country of origin, and date, along with self-reported age and gender of the user when available. We scrubbed the data to remove HIPAA identifiers: names, addresses, IP addresses, phone numbers, Bing user ids, among others. The data were anonymized for Bing for business purposes prior to the researchers’ access. The data sets of malaria and tuberculosis queries were generated in an analogous manner, except with keywords “malaria” as well as “tb” or “tuberculosis.”

Figure 7.1 shows two heat maps for each disease. The top maps illustrate the 2016 disease prevalence (for HIV) or incidence (for malaria/tuberculosis) rates for each country, obtained from the World Bank Databank (WB, 2018). The bottom maps illustrate the fraction of total searches made in each country that contain the specific disease terms. There is a high correlation between the fraction of searches about a given disease in a particular country and the disease rate. The Spearman correlation is $\rho = 0.714$ [0.689, 0.737] for HIV/AIDS,

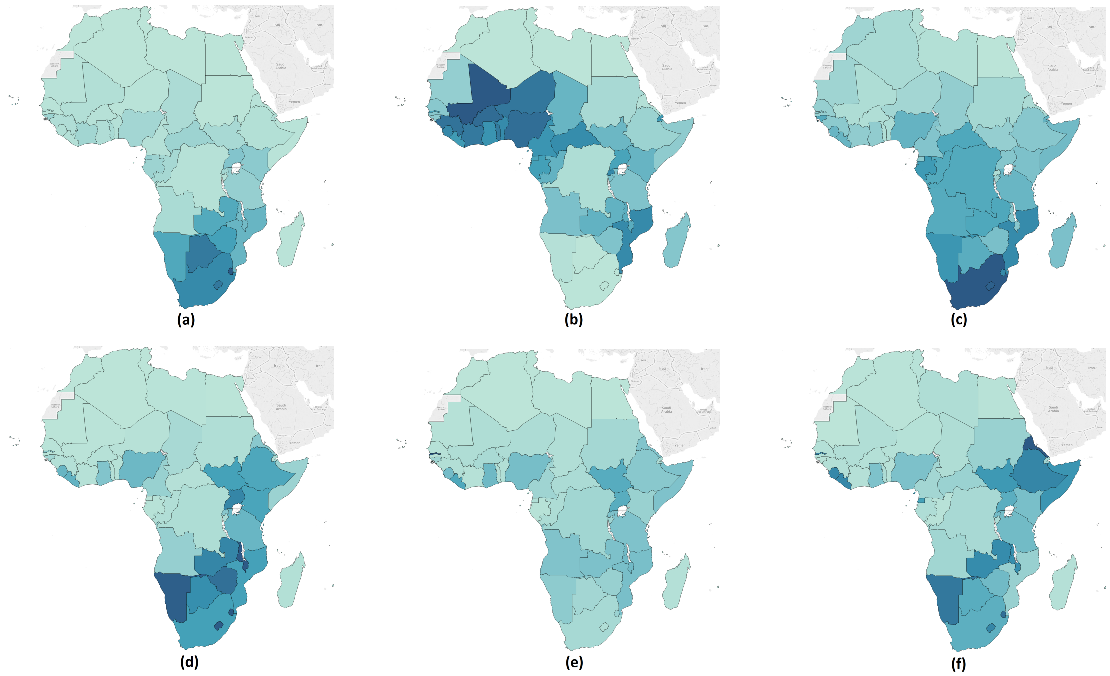


Figure 7.1: Top: Heat maps showing 2016 rates of (a) HIV/AIDS prevalence (ages 15 - 49), (b) malaria incidence, and (c) tuberculosis incidence rates. Bottom: Heat maps showing percentage of total search traffic containing the words (d) “HIV” or “AIDS”, (e) “malaria,” and (f) “tb” or “tuberculosis”.

$\rho = 0.402$ [0.360, 0.442] for malaria, and $\rho = 0.462$ [0.422, 0.499] for tuberculosis. Each correlation coefficient has $p < 0.01$. We view this as reassurance that search queries filtered in this way are pertinent to the diseases in question.

Disease Topics. We extracted topics from each data set using LDA, a standard generative statistical model in which each *document* (in our case, an individual search query) is a distribution over *topics*, and each topic is a distribution over words (Blei et al., 2003). We used the implementation of LDA provided by the Mallet package (McCallum, 2002) and the Differential Language Analysis ToolKit as an interface to Mallet for further analysis (Schwartz et al., 2013, 2017). We retained all default parameters, with the exception of α, the prior on the per-document topic distribution, which we set to 2 since search queries are shorter than the documents for which LDA is typically used.

The number of topics extracted is a free parameter that can be tuned. Choosing a large number of topics leads to the discovery of highly specific topics that overlap in theme, while choosing a small number leads to general, multi-theme, and difficult to interpret topics (Schwartz and Ungar, 2015). Before running our analyses, we ran LDA on each data set with different numbers of topics (10, 20, 50, 100, 200, 500, 1,000, and 2,000). Based on a manual inspection of interpretability and coherence of topics by a health expert, we chose 100 topics for the HIV/AIDS data set and 50 each for the malaria and tuberculosis data sets, which are smaller than the HIV/AIDS data set. These topics were then labeled by a health expert and manually inspected by the authors and checked for any overlooked HIPAA identifiers.

We would ideally like to define representative queries as queries with high weight for the topic. However, the existence of rare words or strings (such as obscure URLs) in a query can result in a query having an artificially high weight for a given topic (abnormally high probability of belonging to a single topic). We thus excluded words that appear fewer than 10 times in the data set. For the same reason, we removed all queries with two or fewer words since these often contained similar issues. (Note that at least one of these must be the name of the disease.)

Additional methods are described below alongside the corresponding results. All statistical significance tests were conducted by correcting for false discovery rate with $\alpha = .05$ of testing either 50 or 100 topics using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995).

7.2 Analysis and Results

Our analyses reveal a rich set of themes. The topics output by LDA range from those about standard health information, such as *Symptoms*, *Drugs*, and *Epidemiology*, to those about hard-to-survey concerns, such as *Stigma* and *Natural Cure*. Table 7.1 shows six sample topics for HIV/AIDS extracted from the data by LDA, which were hand-chosen and labeled by a health expert to illustrate both the breadth of themes that emerged from the analysis as well as the coverage of hard-to-survey topics. We include the topics for malaria and tuberculosis in the appendix. For the HIV/AIDS data set, the full list of topics additionally includes themes such as *Transmission*, *Testing Kits*, *Testing Clinics*, *Gender Inequality*, *Healthy Lifestyle*, *Disease Progression*, and *Celebrity Gossip*. Likewise, the malaria and tuberculosis data sets display a rich set of themes ranging from *Patient Care*, and *Pregnancy*, to *National Programs*.

The second column in Table 7.1 provides a label for each topic along with a measure of the frequency with which the topic occurs in the data. Specifically, given $\theta_{\text{query}} = p(\text{topic}|\text{query})$, the values in parentheses correspond to $\sum_{q \in \text{query}} \theta_q \times p(q)$ over all queries, expressed as a percentage. This corresponds to the popularity of the given topic. Some themes, such as breastfeeding, are captured by more than one topic, so the overall frequency with which these themes occur in the data is higher than the numbers suggest. The third column presents the 20 most representative words, according to the posterior probabilities of word given topic. The final column shows randomly selected queries from among the 100 most closely related to the topic. We show a random sample of queries since the top few most highly ranked queries often differ by only one letter or word. All typos in the queries are unaltered.

Topic	20 Most Representative Words	Sample Queries from Top 100
Symptoms (2.28%)	pain, sign, lymph, swollen, nodes, sore, symptom, symptoms, throat, infection, body, back, positive, pains, stomach, fever, neck, headache, glands, patient	hiv painfull jaw hiv swollen lymph nodes hiv swollen gland throat joints pain hiv
Natural Cure (0.74%)	cure, oil, black, healing, heal, healed, seed, herbs, natural, cures, moringa, kill, cured, testimonials, coconut, traditional, god, garlic, lemon, aloe	prophet bushiri hiv miracles hiv garlic lemon honey coloidal silver hiv testimonials
Epidemiology (0.59%)	statistics, report, 2015, global, unaids, 2016, united, epidemic, besigye, kizza, children, 2014, progress, 2010, response, nations, nigeria, prevalence, million, sa	unaids global aids report mia khalifa hiv hiv 2030
Drugs (0.85%)	drug, treatment, patients, abuse, therapy, drugs, resistance, antiretroviral, substance, adherence, alcohol, art, failure, spread, leads, patient, relationship, transmission, effect	stanford hiv drug resistance hiv drug therapy resistance virological failure hiv
Breastfeeding (0.66%)	positive, baby, mother, breastfeeding, breast, mothers, child, born, feeding, babies, birth, give, breastfeed, infant, infected, feed, milk, pregnant, safe, exposed	hiv exclusive fomular feeding exclusive breast feeding and hiv hiv mom can breast feed baby
Stigma (0.46%)	stigma, issues, discrimination, related, ethical, legal, prevention, safety, pdf, workplace, relating, precaution, work, dies, surrounding, universal, reduce, address	hiv aids ethical dilema safty issues relating to hiv-aids aids stigma in garissa

Table 7.1: Sample LDA topics for the HIV/AIDS data set with representative words and sample queries.

7.2.1 Topic Prevalence by Region and Demographics

We explore whether health information needs, manifested as search queries, vary by country and user demographics. Due to space limitations, we only present results for queries related to HIV/AIDS. The corresponding results for malaria and tuberculosis can be found in the appendix.

For each of the six HIV/AIDS topics listed in Table 7.1, we estimate the number of times an HIV/AIDS topic is queried relative to the overall number of queries for HIV/AIDS for the country. We call this quantity topic prevalence and denote it by prevalence(topic|country), which is a measure of the frequency with which a topic is mentioned in queries from a given country. To estimate the prevalence, we need two values—the frequency the topic is searched for in the country and the overall number of searches for any topics in the country. We do not have the exact number of times a topic is mentioned, but we can estimate the frequency with which the topic is used by utilizing the words associated with a topic using the posterior probabilities, derived from LDA, of a topic given a word, $p(\text{topic}|\text{word})$. We then combine these estimated counts with the relative frequencies of a word given a country, frequency(word|country) to get the overall prevalence of the topic in the country:

$$\sum_{\text{word}\in\text{country}} p(\text{topic}|\text{word}) \times \text{frequency}(\text{word}|\text{country})$$

Here, frequency(word|country) is derived from maximum likelihood estimation given all words used in queries from the country: the ratio of the number of times a word appears and the total count of all words for the country.[2]

[2]Since we use an estimate of the number of times a topic is searched for, we provide results for estimating frequency differently, namely by considering only the top 50 words associated with a topic ranked by $p(\text{topic}|\text{word})$, in the appendix. We obtain qualitatively similar results

Topic Prevalence by Country. To explore the association of topic prevalence with the adult HIV prevalence rates across countries, we ran a linear regression using the prevalence values for each of the 100 topics within a given country as the explanatory variables and the 2016 HIV prevalence rate in that country as the dependent variable. (Note that throughout this text, we consider the adult HIV prevalence rate, which is set to ages 15-48, but simply refer to it as the HIV prevalence rate.) Of the six topics listed in Table 7.1, we found a significant relationship between the *Stigma* topic and the HIV prevalence rate (r = 0.473; multi-test corrected $p < 0.01$), as illustrated in Figure 7.2.

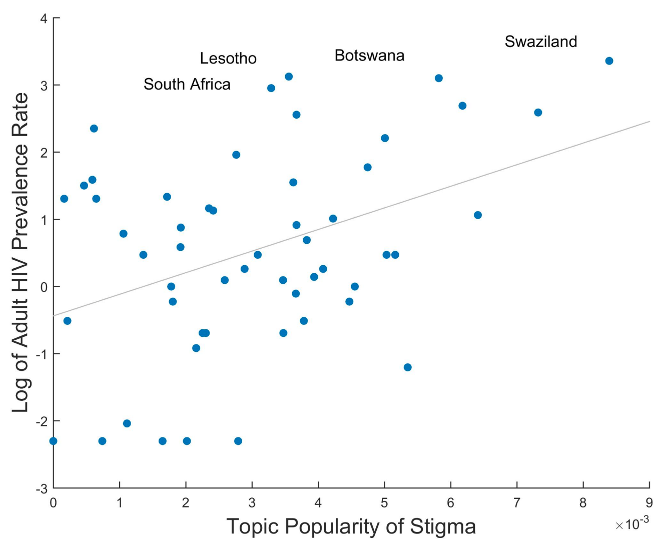


Figure 7.2: Comparison between topic popularity of *Stigma* in each country with the log of the 2016 adult HIV prevalence rate. Countries with higher topic popularity for *Stigma* tend to have higher HIV prevalence rates.

The observation that the popularity of the *Stigma* topic is correlated with HIV prevalence is consistent with findings from the public health literature. In particular, smaller-scale studies (often based on survey data), have shown that HIV-related stigma can lead to more risky behavior, lower testing rates, and de-

when we consider all the words or the top words to estimate the frequency the topic is searched for in a country.

creased adherence to antiretroviral therapy, all of which increase transmission rates (Chan and Tsai, 2015; Genberg et al., 2009). An outlier in the above figure, *Stigma* is nearly twice as frequent in Botswana as in Lesotho, despite the two countries having comparable HIV prevalence rates (22.2% versus 22.7%). This is consistent with literature indicating that Botswana has struggled with discrimination issues including proposals for mandatory HIV testing (Sarumi and Strode, 2015). Similarly, Swaziland is also known to have one of the highest rates of HIV prevalence, as well as stigma surrounding the disease (Root, 2010). Further, despite high HIV prevalence rates, the level of stigma has been declining in South Africa, in part due to targeted HIV destigmatization efforts (Mbonu et al., 2009).

Other topics also exhibit variance in popularity by country. To explore this, we ran a logistic regression for each country, with the standardized topic weights (distribution) of a given search query as the explanatory variables and a binary dependent variable indicating whether or not the query originated in that country. There is a notable contrast in the popularity of queries associated with the *Natural Cure* topic across the countries. For instance, the *Natural Cure* topic is popular in Malawi ($\beta = 0.05; p < 0.05$), which has a relatively high HIV prevalence rate of 10.6, and in Botswana ($\beta = 0.07; p < 0.01$), which has an HIV prevalence rate of 22.2%. In Mozambique, which has an HIV prevalence rate of 11.5%, the popularity of the *Natural Cure* topic is relatively low ($\beta = -0.07; p < 0.01$).

Topic Prevalence by User Demographics. To explore topic popularity by gender, we ran a logistic regression with the topic weights of a given search query as the explanatory variables and the self-reported gender of the user as the

Ages 18–24 *(0.083)*: symptoms, signs, early, women, men, infection, stages, months, symptoms, earliest, children, major, syptoms, systoms, rare
Ages 25–34 *(0.070)*: positive, negative, partner, person, man, sex, woman, tested, im, infected, pregnant, baby, husband, wife, infect
Ages 35–49 *(0.063)*: cure, latest, news, research, treatment, discovery, today, vaccine, update, 2015, 2016, 2017, google, breakthrough, recent

Women *(0.115)*: positive, baby, mother, breastfeeding, breast, mothers, child, born, feeding, babies, birth, given, breastfeed, infant, infected
Men *(0.133)*: cure, news, 2016, latest, vaccine, breaking, 2017, www, development, today, headline, updates, breakthrough, found, feb

Table 7.2: Popular topics by age and gender for the HIV/AIDS data set.

dependent variable, limiting ourselves to those queries for which user demographic information was available. Similarly, to explore topic popularity by age group, we ran an ordinary least squares linear regression, again with the topic weights of a given search query as the explanatory variables, but now with users' self-reported age group as the dependent variable. We then ordered the topics by their respective correlation coefficients. The fifteen words with the highest weight from the top ranked topic for each group are shown in Table 7.2.

Our analysis reveals that topics related to news on HIV/AIDS cures are more popular among men, as well as the 35–49 age group. Topics related to breastfeeding, pregnancy, and family care are more popular among women. For the 18–24 and 25–34 age groups, topics related to symptoms are more popular. Among the former group, topics related to the socioeconomic implications of HIV/AIDS, such as gender inequality, are more popular, while topics related to concerns about transmission to partner and child are more popular among the 25–34 age group.[3]

[3]Some themes appear in more than one topic. When we identify a novel pattern relating a particular topic to demographics or location in the data (e.g., "breastfeeding is more popular among women"), we confirmed the same pattern exists for the most similar topics. To measure

	Women	Ages 18–24	Ages 25–34	Ages 35–49
Symptoms	-0.052**	0.000	-0.019*	-0.018*
Natural Cure	-0.010	-0.050**	-0.018*	0.043**
Epidemiology	-0.052**	-0.080**	-0.019*	0.019*
Drugs	-0.016	-0.020**	-0.041**	0.030**
Breastfeeding	0.115**	-0.031**	0.061**	-0.008
Stigma	0.025**	0.032**	-0.047**	0.004

Table 7.3: Topic popularity by user demographics for the HIV/AIDS topic.

Finally, we looked at the topic popularity of the six topics of interest from Table 7.1. Table 7.3 lists the correlation coefficients, where $**$ indicates a p-value of less than 0.01 and $*$ indicates a p-value of less than 0.05.

We again confirm *Breastfeeding* has a higher correlation coefficient for women than for men and for the 25–34 age group compared to the other age groups. Less expected, women and users aged 18–24 are more interested in *Stigma* compared to their demographic counterparts. *Natural Cure* has the highest popularity among the oldest age group (35–49), and the lowest among the youngest age group (18–24). Despite expressing higher interest in *Natural Cure*, the 35–49 age group also has more interest in *Drugs* compared to the other age groups.

7.2.2 User Behavior and Quality of Results

We examine whether user behavior and the quality of search results returned vary across different topics. Differences here would highlight unmet health information needs, concentration of misinformation related to specific topics, and differences in user satisfaction by topic.

similarity between topics, we calculate the pairwise hamming distance between topics using the top 20 most representative words. If the observed pattern does not hold across similar topics, we omit the pattern from our positive findings.

We used an expanded version of our HIV/AIDS data set consisting of only those queries that were made during June 2017. In addition to raw queries, country, and search date, this data set contains a list of the first 10 organic web pages returned to the user for each query. It also contains information about which web pages the user clicked on, the amount of time spent on each web page, and the total time spent on the *results page*, the page containing the ten initial links presented to a user after entering a search query.

To compare user behavior across topics, we focused on several standard metrics from the information retrieval literature (Manning et al., 2010). *Dwell time* measures the total amount of time a user spends looking at the results page and any links that are followed. *Click count* is the total number of links on which a user clicks.[4] Note that these metrics can be used to measure various properties related to user engagement and satisfaction. For instance, dwell time can be used to measure both interest in a web page and ease-of-use, depending on the context and intent of the search query. In our study, we use these metrics to measure user activity. Specifically, we are interested in measuring whether there is variance in user activity by topic as measured by these metrics.

Plots (a) and (b) in Figure 7.3 show how these metrics vary by topic. Both dwell time and click count are significantly lower for the *Natural Cure* topic compared with the other topics of interest. That is, on average, users issuing queries related to the *Natural Cure* topic spend less time exploring the results page and click on fewer links. There are many reasons why this may be the case. It could simply be selection bias—perhaps different types of users search

[4]We also examined *successful click count*, which measures the number of pages a user clicks on with dwell time at least 30 seconds, and *maximum dwell time*, which measures the maximum amount of time spent on a web page. Results for these are similar to the click count and dwell time results, respectively.

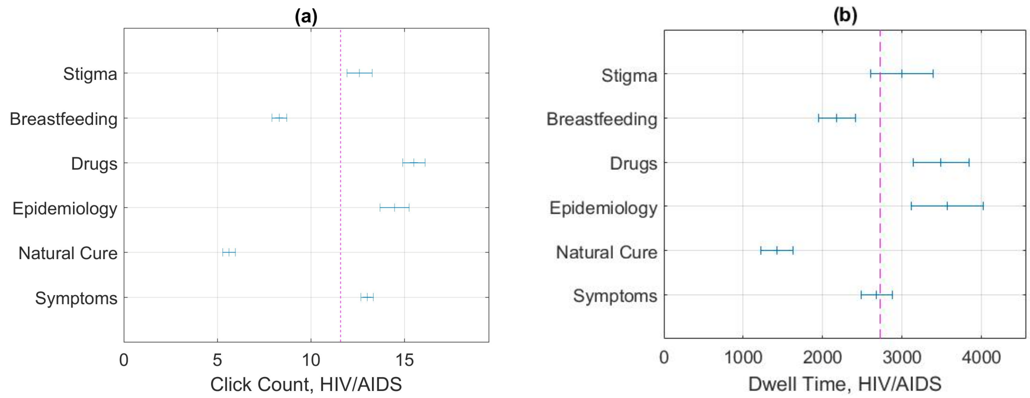


Figure 7.3: Average (a) click count and (b) total dwell time for queries associated with HIV/AIDS topics of interest. The vertical line represents the mean value across topics.

for queries related to the *Natural Cure* topic compared with *Epidemiology*, *Drugs*, or other topics. It could be that users seeking information related to the *Natural Cure* topic find the information they are seeking faster. Another possibility is that the quality of information returned could vary by query.

To examine variance in the quality of content returned, for each of the topics in Table 7.1, we extracted the first link returned to the user at the top of the web results page for each of the 30 queries most strongly associated with the topic. We consider only distinct user/query pairs, which means we ignored duplicate queries from the same user. Each resulting link was independently evaluated for quality (described in terms of relevance, accuracy, and objectiveness, as is standard in information retrieval (Hasan and Abuelrub, 2011)) and was ranked by three research assistants on a scale of 1 to 5, with higher values indicating better quality. The research assistants who provided rankings have graduate-level training in medicine or public health, and each website was evaluated by at least one research assistant specializing in the disease of interest. We took the average of these three ratings as the rating for a web page.

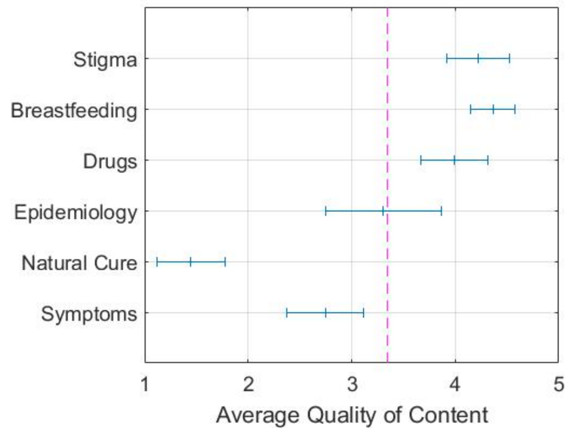


Figure 7.4: Average quality of content of web pages returned to users for the 30 queries most strongly associated with each HIV/AIDS topics of interest.

Figure 7.4 shows the average rating across all links and all raters for each topic. On average, the quality of links returned for queries related to the *Natural Cure* topic is low, with an average quality rating of 1.45. In contrast, links returned for queries related to the *Stigma*, *Breastfeeding*, and *Drugs* topics have much higher average quality ratings (4.22, 4.36, and 3.99, respectively). A t-test comparison between *Natural Cure* with each of these topics yields $p < 0.01$. Results for malaria and tuberculosis are similar. Consistent with prior research on H1N1 outbreaks (Hill et al., 2011), the quality of content that is returned to users varies by topic, especially when we compare *Natural Cure* vs. *Drugs*. The figures below show partial result pages shown based on real and popular search queries from users. Note that the discrepancy in the average quality of results persists across the Google search engine as well.

Our analysis could also be used to investigate the quality and volume of information available to individuals with different health information needs. To get a sense for how much high-quality public health information on natural cures for HIV/AIDS is available, we used the queries most highly associated

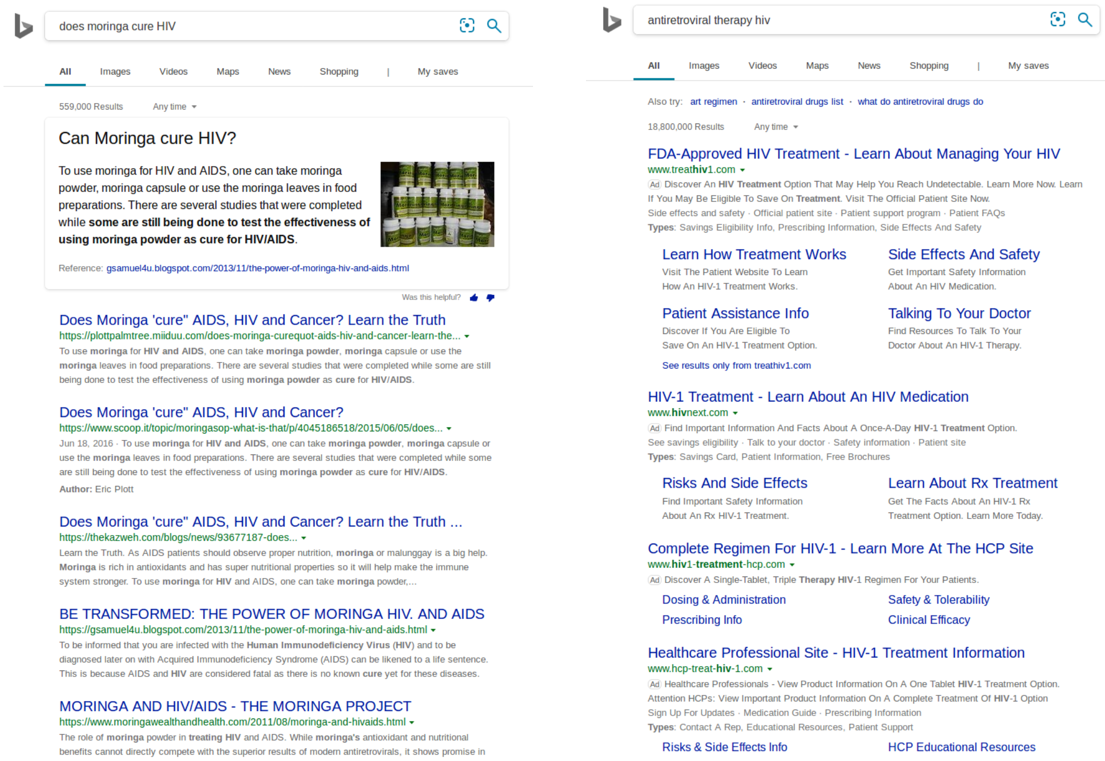


Figure 7.5: Sample result pages shown for two real search queries on the Bing search engine.

with the *Natural Cure* and *Drugs* topics on Bing to search authoritative websites on HIV/AIDS. The authoritative websites include the World Health Organization (WHO), the Joint United Nations Programme on HIV/AIDS (UNAIDS), the Center for Disease Control and Prevention (CDC), and the National Institutes of Health (NIH). We posed each of the top 30 queries for *Natural Cure* in turn to each of the aforementioned authoritative websites and noted the number of web pages that were returned on each website for each query. We performed the same actions for the *Drugs* topic. We then reported the average number of web pages available for the 30 queries corresponding to the two topics by website.

We found that on the CDC site, there were an average of 56,705.85 web pages corresponding to *Natural Cure* (over the 30 queries corresponding to the topic) compared to 258,948.6 for the top 30 queries for *Drugs* ($p = 0.01$). Similarly, for the NIH site, there were $91,840.8$ and $456,982.3$ ($p = 0.00$); for the WHO, there were $46,600.1$ and $305,528.2$ ($p = 0.00$); and for UNAIDS, there were $8,926.5$ and $65,954.0$ ($p = 0.02$) web pages for *Natural Cure* and *Drugs*, respectively. By this measure, there are consistently fewer high-quality documents for natural cures than for pharmaceutical drugs on authoritative websites.

7.3 Discussion

In this chapter, we have shown that search data, with well-chosen analyses, can provide valuable insights into the health information needs, concerns, and misconceptions of individuals across Africa. Such analyses can complement existing top-down approaches and allow us to narrow the gap in available health data between developing and developed nations. We conclude with a discussion of the limitations of our techniques as well as implications and next steps for future work.

Limitations. There are several limitations to using search data. First, the Bing users we study – and Internet users in general – are not a representative sample of the entire population of Africa, as is the case with most work reliant on large Web- and social-media based data. Throughout this work, we have included analyses which give evidence that the data is associated with the diseases at hand. However, since ground-truth data, and especially disaggregated by demographics, is hard to come by due to the health data gap, we are further limited in our ability to measure the representativeness of this data. It is therefore challenging to extrapolate observations obtained through the analysis of search data to the wider population of countries, and the health concerns of entire communities who are not on the web could be overlooked. Still, many findings corroborated previous smaller scale studies using representative samples. This concern will diminish as Internet penetration continues to grow, but the studied population will never be fully representative. A second limitation is that the results of this study depend on proprietary data from Bing which can limit the ability for health organizations to extend the research.

Another limitation is the use of imprecise language in search queries, as well as queries in languages other than English. An initial exploration of the Bing query logs showed that many users search for HIV/AIDS, malaria, and tuberculosis by their English names, but it is likely that the filtering method we used still led us to exclude many relevant searches. Furthermore, the excluded searches are more likely to come from regions in which the use of English names is less common, further biasing the data collection. These concerns could be amplified for other illnesses, such as respiratory infections, which have multiple common names in different languages and for which users commonly search for symptoms instead of the disease name itself. A multi-language approach would be necessary to more fully extract all of the information on individuals' health information needs that is captured in search data.

Practical Implications. Our methods have great potential to inform targeted education efforts in data-sparse regions. Gender and age impact an individual's chance of contracting HIV/AIDS, malaria, or tuberculosis (Germain, 2009), and health information needs are often specific to demographic groups and geographic locations. For these reasons, stakeholders have emphasized the need for gender-responsive and age-responsive programming in resource-constrained regions. Efforts to understand health information needs in developing nations by demographic group have mostly used surveys and interviews, which are limited in their scale (Abimanyi-Ochom et al., 2017; Li et al., 2004; Wang et al., 2008). Search data allows us to study health needs at a much larger scale.

Search engines themselves could potentially be an effective platform for implementing targeted interventions to improve access to health information. For instance, gender- or age-specific targeted advertisements for health campaigns could be triggered by queries associated with specific health topics. Insights garnered from the analysis of search data could help health organizations prepare material and develop interventions aimed at regions where specific misconceptions are especially common. These interventions could take the form of highlighted high-authority links to discourage misinformation, advertisements for support groups triggered by searches related to stigma, or advertisements for testing clinics for testing-related searches. Recent work has studied the potential to use computational techniques to combat health misinformation on social media (Ghenai and Mejova, 2018).

Search data could also be used to monitor other aspects of public health, for example by providing marketing surveillance for new medications or measuring the impact of public health campaigns. In principle, search engines and health organizations could also work together on case finding, a strategy that directs resources at individuals or groups suspected to be at risk for a particular disease, which is a key strategy in communicable disease outbreak management. Of course, this pursuit would need to be handled with great care, with consideration for the risks and ethics involved.

Finally, as it is detailed and available in real time, search data could be especially valuable for monitoring the impacts of emerging health concerns in developing nations. For instance, noncommunicable diseases such as cancer, cardiovascular disease, and diabetes are of growing concern due to the expansion of the middle class in developing countries and a lack of resources and programs aimed at minimizing their impact (Sambo, 2014). Since the portion of the population affected by these diseases is likely to have Internet access, search data could play an instrumental role in understanding attitudes about these diseases, implementing interventions to improve access to health information, and highlighting overlooked aspects of the impacts of these diseases.

CHAPTER 8

OVERCOMING CHALLENGES IN WORD-OF-MOUTH RECOMMENDATIONS

A rich line of work studies the effectiveness of spreading ideas and innovations based on person-to-person recommendation within a social network — a process often termed *word-of-mouth recommendations* (Jurvetson, 2000) and closely connected to the broader sociological literature on the *diffusion of innovations* in social networks (Rogers, 1995). A key genre of theoretical question that emerged early in this literature is the problem of optimally "seeding" an innovation – such as relevant information or a product – in a social network through the selection of a set of initial adopters (Domingos and Richardson, 2001; Kempe et al., 2003; Richardson and Domingos, 2002). This strategy has been leveraged in a variety of contexts, ranging from health campaigns to viral marketing. Within health, campaign strategies relying on word-of-mouth recommendations have demonstrated efficacy in contexts including vaccinations, West Nile virus education, healthy life-styles, health behavior related to HIV, and many other examples (Averett et al., 2005; Compton and Pfau, 2009; Fox et al., 2006; Gelb and Johnson, 1995; Zimbardo and Leippe, 1991).

In word-of-mouth recommendations, a designer – such as a health authority or a firm – that has an innovation, whether in the form of information or a product, aims to see this innovation adopted by a group of agents on a social network. It is often the case that the designer cannot target all the participants in the network, and so they seek to target the most influential ones so as to maximize exposure and create a cascade of adoptions. Approaches to this question have generally been based on objective functions in which the goal is to maximize the number of people who are reached by the network cascade — or more generally, in which the objective function monotonically increases in the number of people reached. In this chapter, we explore a question motivated by health campaigns: how does one diffuse innovation over a network when there may be agents who are not susceptible to the innovation and whose exposure to this innovation may be costly to the designer.

The dangers of overexposure. Research in both public health campaigns and marketing has provided diverse evidence that the benefits of a word-of-mouth recommendations are not in fact purely increasing in the number of people reached. An influential example of such a finding is the *Groupon effect*, in which viral marketing via Groupon coupons led to lower Yelp ratings (Byers et al., 2012). This research notes the negative effect Groupon had on average Yelp ratings and provides arguments for the underlying mechanism; one of their central hypotheses is that by using Groupon as a matchmaker, businesses may be attracting customers from a portion of the population that is less inclined to like the product. In another example, Kovcs and Sharkey (2014) discuss a setting on Goodreads where books that won prestigious awards (or are short-listed for them) attracted more readers following the announcement, which again led to a drop in the average rating of the book on the platform. Aizen et al. (2004) show a similar effect for on-line videos and other media; they received a discontinuous drop in their ratings when a popular blog linked to them, driving users to the item who may not be interested in it. Research in marketing has shown that exposure to different groups and influence between such groups can help or hurt adoption (Berger and Heath, 2007; Hu and den Bulte, 2014; Joshi et al., 2016). For example, Hu and den Bulte (2014) argue that agents adopt products to boost their status; and so as word-of-mouth effects for a product become stronger among middle-income individuals, there might be a negative impact on adoption among higher-income individuals. Similar behavior is observed in health campaigns. For instance, research on anti-smoking or drugs campaigns such as DARE has shown that targeting information to populations that are not receptive to it can have negligible effects and might even have a negative impact, costing the designer both in the expenses of targeting and achieving the opposite of the desired outcome (Hamilton, 1997). Indeed, researchers in public health have long discussed the importance of segmenting populations and exposing groups to anti-smoking campaigns whose themes the group is most susceptible to in order to maximize the impact of future campaigns (Wakefield et al., 2003).

We think of these effects collectively as different forms of *overexposure*; while reaching many potential agents on a network is not a concern in and of itself, the empirical research above suggests that there may exist particular subsets of the population — potentially large subsets — who will react negatively to the innovation. When a word-of-mouth cascade reaches members of this negatively inclined subset, the campaign can suffer negative payoff that may offset the benefits it has received from other parts of the population. This negative payoff can come in the form of harm to the designer's reputation, perhaps through latent population impressions and efficacy of future campaigns, (Cravens et al., 2003; Cretu and Brodie, 2007).

Despite the importance of these considerations, this overexposure phenomena has not been incorporated into models of influence-based word-of-mouth recommendations in social networks. In this chapter, we ask what types of algorithmic issues arise when we seek to spread a word-of-mouth cascade through a network, but must simultaneously ensure that it reaches the "right" part of the audience — the potential individuals who will like the innovation, rather than those who will react negatively to it?

Our Results

In this chapter, we propose a basic theoretical model for the problem of seeding a cascade when there are benefits from reaching positively inclined individuals and costs from reaching negatively inclined individuals. Note that while this model, and the set of questions considered here, are inspired by health campaigns and marketing, they may be applicable in a broad range of domains where word-of-mouth recommendations may be an effective means for diffusing innovations.

There are many potential factors that play a role in the distinction between positively and negatively inclined individuals, and for our model we focus on a stylized framework in which each innovation has a known parameter ϕ in the interval $[0, 1]$ that serves as some measure for the breadth of its appeal. At this level of generality, this parameter could serve as a proxy for a number of things, including quality, a normalized measure for the likelihood that an individual will find the information compelling, or – in the case where the social network represents a population defined by a specific interest – compatibility with the core interests of network's members.

Each node in the network is an *agent* who will evaluate the innovation when they first learn of it; agents differ in how *critical* they are of new innovations, with agents of low criticality tending to like a wider range of innovations and agents of high criticality tending to reject more innovations. Thus, each agent i

has a criticality parameter θ_i in the interval $[0, 1]$; since we assume that the designer has a history of spreading innovations to this network over a period of time, it knows this parameter θ_i for each agent i. When exposed to an innovation, an agent accepts the innovation if $\phi \geq \theta_i$ and announces the innovation to their neighbors, leading to the potential for a cascade. However, if $\phi < \theta_i$, then the agent rejects the innovation, which results in a negative payoff to the designer; the cascade stops at such agents i, since they do not announce it to their neighbors.

The designer's goal is to seed the innovation to a subset of the nodes in the network — the *seed set* — resulting in a potential cascade of further nodes who learn about the innovation, so as to maximize its overall payoff. This payoff includes a positive term for each agent i who sees the innovation and accepts it since $\phi \geq \theta_i$, and a negative term for each agent i who sees the innovation and rejects it since $\phi < \theta_i$; agents who are never reached by the cascade never find out about the innovation, and the designer gets zero payoff from them.

We obtain theoretical results for two main settings of this problem: the *unbudgeted case*, in which the designer can initially announce the innovation to an arbitrary seed set of nodes, and the *budgeted case*, in which the designer can seed the project to at most k nodes, for a given parameter k. We note that typically in influence maximization problems, the unbudgeted case is not interesting: if the payoff is monotonically increasing in the number of nodes who are exposed to the innovation, then the optimal unbudgeted strategy is simply to show the innovation to everyone. In a world with negative payoffs from overexposure, however, the unbudgeted optimization problem becomes non-trivial: we must tradeoff the benefits of showing the innovation to individuals who will like it against the negatives that arise when these individuals in turn share it with others who do not.

For the unbudgeted problem, we give a polynomial-time algorithm for finding the optimal seed set. The algorithm uses network flow techniques on a graph derived from the underlying social network with the given set of parameters θ_i. In contrast, we provide an NP-hardness result for the budgeted problem.

We then provide a natural generalization of the model: rather than each agent exhibiting only two possible behaviors (rejecting the innovation, or accepting it and promoting it), we allow for a wider range of agent behaviors. In particular, we will assume each agent has three parameters which control whether the agent ignores the innovation, views but rejects the innovation, accepts the innovation but does not broadcast it to its neighbors, and accepts the innovation and announces it to neighbors. We show how to extend our results to this more general case, obtaining a polynomial-time algorithm for the unbudgeted case and an NP-hardness result for the budgeted case.

Finally, we perform computational simulations of our algorithm for the unbudgeted case on sample network topologies derived from moderately-sized social networks. We find an interesting effect in which the performance of the optimal algorithm transitions between two behaviors as ϕ varies. For small ϕ the payoff grows slowly while a baseline that promotes the innovation to every agent i with $\theta_i < \phi$ achieves negative payoff (reflecting the consequences of overexposure). Then, for large ϕ, the payoff grows quickly, approaching a simple upper bound consisting of all i for which $\theta_i < \phi$.

8.1 Preliminaries

There is an innovation with a parameter $\phi \in [0, 1]$, measuring the breadth of its appeal. G is an unweighted, undirected graph with n agents as its nodes. For each agent i, the agent's criticality parameter $\theta_i \in [0, 1]$ measures the minimum threshold for ϕ the agent demands before adoption. Thus, higher values of θ_i correspond to more critical agents. We assume that these values are fixed and known to the designer.

The designer chooses an initial set of agents $S \subseteq V$ to "seed" with the innovation. If an agent i sees the innovation, it adopts it if $\theta_i \leq \phi$ and rejects it if $\theta_i > \phi$. We say that an agent i is *accepting* in the former case and *rejecting* in the latter case. Each accepting agent who is exposed to the innovation announces it to their neighbors, who then, recursively, are also exposed to the innovation. We will assume throughout that the designer chooses a seed set consisting entirely of accepting nodes (noting, of course, that rejecting nodes might subsequently be exposed to the innovation after nodes in the seed set announce it to their neighbors).

We write $V(S)$ for the set of agents exposed to the innovation if the seed set is S. Formally, $V(S)$ is the set of all agents i who have a path to some node in $j \in S$ such that all of the internal nodes on the i-j path are accepting agents; this is the "chain of recommendations" by which the innovation reached i. Among the nodes in $V(S)$, we define $V^+(S)$ to be the set of agents who accept the innovation and $V^-(S)$ to be the set of agents who reject the innovation.

The payoff function associated with seed set S is

$$\pi(S) = |V^+(S)| - |V^-(S)|. \tag{8.1}$$

We can, more generally, assume that there is a payoff of p to accepting the innovation and a negative payoff of q to rejecting the innovation, and we set the payoff function to be

$$\pi(S) = p|V^+(S)| - q|V^-(S)|. \tag{8.2}$$

We will call this the *generalized payoff function*, and simply refer to Equation 8.1 as the *payoff function*. The overarching question then is:

Problem 30. Given a set of agents V with criticality parameters θ_i ($i \in V$) on a social network $G = (V, E)$, and given an innovation of quality ϕ, what is the optimal seed set $S \subseteq V$ that the designer should target in order to maximize the payoff given by Equation (8.2)?

In contrast to much of the influence maximization literature, we assume that the agents' likelihood of adoption, once exposed to this innovation, is not affected by which of their neighbors have accepted or rejected the innovation. This differs from, for instance, models in which each agent requires a certain fraction (or number) of its neighbors to have accepted the innovation before it does; or models where probabilistic contagion takes place across the edges. These all form interesting directions for further work; here, however, we focus on questions in which the intrinsic appeal of the innovation, via ϕ, determines adoption decisions, and the social network provides communication pathways for other agents to hear about the innovation.

Before proceeding to the main result, we develop some further terminology that will be helpful in reasoning about the seed sets.

Definition 31. Let i be an accepting node, and let $S = \{i\}$. Then we say that $V(S)$ is the *cluster* of i, denoted by C_i; we call $V^+(S)$ the *interior* of C_i and denote it by C_i^o, and we call $V^-(S)$ the *boundary* of C_i and denote it by C_i^b.

We denote the payoff corresponding to the seed set $S = \{i\}$ by π_i. Note that,

$$\pi_i = p|C_i^o| - q|C_i^b|.$$

Lemma 32. *Given an accepting node $i \in V$, and a node $j \in C_i^o$, we have $C_i = C_j$.*

Proof. If j is in the interior of C_i, then there exists a path $(k_1, k_2, \cdots, k_\ell)$ in G, where $i = k_1$ and $j = k_\ell$ such that each node along the path has $\theta \leq \phi$. That is, each node k_i is exposed to and accepts the innovation as a result of k_{i-1}'s advertisement. We would like to prove that if $S = \{j\}$, then i would be exposed to the innovation. Equivalently, we want to show there exists a path from j to i of nodes with $\theta \leq \phi$; but this is precisely the path $(k_\ell, k_{\ell-1}, \cdots, k_1)$. □

For an arbitrary seed set S, the set $V(S)$ may consist of multiple interior-disjoint clusters, which we label by $\{C_1, C_2, \cdots, C_k\}$, where $k \leq |S|$. Note that each of these clusters might be associated with more than one agent in the seed set and that $\cup_{i=1}^k C_i = V(S)$. (Likewise, $\cup_{i=1}^k C_i^o = V^+(S)$ and $\cup_{i=1}^k C_i^b = V^-(S)$.)

Given a seed set S and corresponding clusters, a direct consequence of Lemma 32 is that adding more nodes already contained in these clusters to the seed set does not change the payoff.

Lemma 33. *Given a set S' of accepting nodes such that $S \subseteq S' \subseteq V(S)$, we have $\pi(S) = \pi(S')$.*

It therefore suffices to seed a single agent within a cluster. Given a cluster C_i, we will simply pick an arbitrary node in the interior of the cluster to be the canonical node i and use that to refer to the cluster even if C_i is formed as a result of seeding another node $j \in C_i^o$.

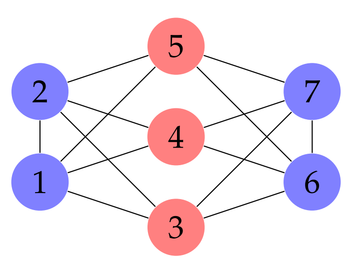
8.2 Modeling Overexposure in Networks

In this section, we tackle the optimization question posed earlier. We first tackle a basic setting where agents are only one of two types. We then use this as a building block for a more general model.

8.2.1 Main Model

Given that all θ_i are known to the designer, a naive approach would suggest to seed all i where $\pi_i \geq 0$. While this is guaranteed to give a nonnegative payoff, S need not be optimal.

Example 34. Consider the graph below, where nodes in blue accept the innovation and those in red reject the innovation.



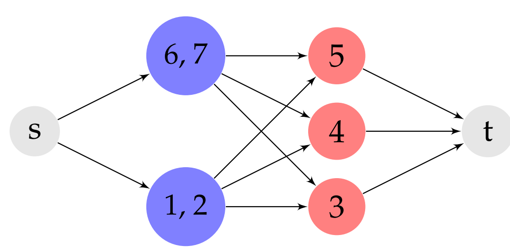
Suppose $p = q = 1$. Then, a naive approach would set $S = \emptyset$, since each of the resulting clusters to the blue nodes has negative payoff. However, setting

$S = \{1, 2, 6, 7\}$ has payoff 1.

This phenomenon is a result of the fact that the clusters C_i might have boundaries that intersect nontrivially. Thus, there could be agents whose $\pi_i < 0$, but C_i^b is in a sense "paid for" by seeding other agents; and hence we could have a net-positive payoff from including i subject to seeding other agents whose cluster boundaries intersect with C_i^b. Using this observation, we will give a polynomial-time algorithm for finding the optimal seed set under the generalized payoff function using a network flow argument. We first begin by constructing a flow network.

Given an instance defined by G and ϕ, we let $\{C_1, C_2, \cdots, C_k\}$ be the set of all distinct clusters in G, with disjoint interiors. We form a flow network as follows: set $A = \{1, 2, \cdots, k\}$ corresponding to the canonical nodes of the clusters above and R be the set of agents in the boundaries of all clusters. We add an edge from the source node s to each node $i \in A$ with capacity $p \cdot |C_i^o|$ and label this value by cap_i, and an edge from each node $j \in R$ to t with capacity q. We add an edge between i and j if and only if $j \in C_i^b$, and set these edges to have infinite capacity. We denote this corresponding flow network by G_N.

Example 35. For the above example, G_N is



In this example, the edges out of s have capacities $2p$ and edges into t have capacities q. Edges between blue and red notes have infinite capacity. Assuming $4p \geq 3q$, the min-cut (X, Y) has $Y = \{t\}$ and all other nodes in X.

Lemma 36. *Given a min-cut (X, Y) in G_N, the optimal seed set in G is $A \cap X$.*

Proof. The min-cut (X, Y) must have value at most $q|R|$ since we can trivially obtain that by setting the cut to be $(V(G_N)\backslash\{t\}, \{t\})$. Given a node $i \in X$, which we recall corresponds to the canonical node of a cluster, if a node $j \in R$ is exposed to the innovation as a result of seeding any node in the cluster, then $j \in X$. Otherwise, we would have edge (i, j) included in the cut, which has infinite capacity contradicting the minimality of the cut (X, Y). Therefore, the min-cut will include all nodes in the seed set as well as all nodes that are exposed to the innovation as a result of the corresponding seed set in S.

Note that the edges across the cut are of two forms: (s, i) or (j, t), where $i \in A$ and $j \in R$. The first set of edges contribute $\sum_{i \in A \cap Y} \text{cap}_i$ (recall $\text{cap}_i = p \cdot |C_i^o|$) and the latter contributes $|R \cap X|q$. Therefore, the objective for finding a min-cut can be equally stated as minimizing,

$$\sum_{i \in A \cap Y} \text{cap}_i + |R \cap X|q.$$

over cuts (X, Y). Note that

$$\sum_{i \in A} \text{cap}_i = \sum_{i \in A \cap X} \text{cap}_i + \sum_{i \in A \cap Y} \text{cap}_i \, .$$

Therefore, we have:

$$\begin{aligned}
&\left(\sum_{i \in A \cap Y} \text{cap}_i + |R \cap X|q\right) \\
&= \left(\sum_{i \in A} \text{cap}_i - \sum_{i \in A \cap X} \text{cap}_i + |R \cap X|q\right) \\
&= \sum_{i \in A} \text{cap}_i - \left(\sum_{i \in A \cap X} \text{cap}_i - |R \cap X|q\right)
\end{aligned}$$

Note the term $(\sum_{i \in A \cap X} \text{cap}_i - |R \cap X|q)$ is precisely what the payoff objective function is maximizing, giving a correspondence between the min-cut and optimal seed set. $\square$

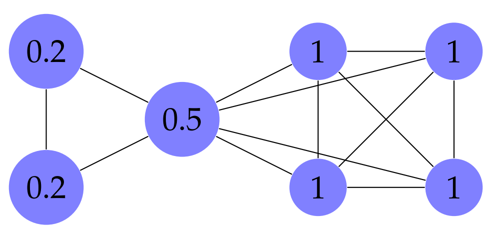
We therefore have the main result of this section:

Theorem 37. *There is a polynomial-time algorithm for computing the optimal seed set for Problem 30 when there are no budgets for the size of the seed set.*

We can adapt the results above to the directed graph case analogously. Here, we let A be the set of nodes i for which $\theta_i \geq \phi$ and set the capacity of edges (s, i) with $i \in A$ to be p. We add an edge (i, j) with infinite capacity if there exists a directed walk from j to i. As above, we set the capacity of edges (j, t) to be q.

An interesting phenomenon is that the payoff is not monotone in ϕ even when considering optimal seed sets. Take the following example:

Example 38. Suppose we are given the network below with the numbers specifying the θ_i for each corresponding node:



If $\phi \in [0, 0.2)$, we cannot do better than the empty-set. If $\phi \in [0.2, 0.5)$, the seed set that includes either of the two leftmost nodes gives a payoff of 1, which is optimal. For the case where $\phi \in (0.5, 1)$, the empty-set is again optimal.

This example gives a concrete way to think about overexposure phenomena.

8.2.2 Generalized Model

We now consider the generalized model where there are three parameters corresponding to each agent i, $\tau_i \leq \theta_i \leq \sigma_i$. An agent considers an innovation if $\tau_i \leq \phi$, adopts an innovation if $\theta_i \leq \phi$, and announces it to their friends if $\sigma_i \leq \phi$. If an agent is exposed to an innovation but $\phi < \tau_i$, then the payoff associated with the agent is 0. If $\phi \in [\tau_i, \theta_i)$, then the agent rejects the innovation, for a payoff of $-q < 0$. As before, there is a payoff of $p > 0$ if the agent accepts the innovation; however, the agent only announces the innovation to its neighbors after accepting if $\phi \geq \sigma_i$. We therefore have four types of agents:

- Type I: Agents for which $\phi < \tau_i$,
- Type II: Agents for which $\phi \in [\tau_i, \theta_i)$,
- Type III: Agents for which $\phi \in [\theta_i, \sigma_i)$, and
- Type IV: Agents for which $\phi \geq \sigma_i$.

We denote the set of all agents of Type I by T_1 (and likewise for the other types). The basic model above is the special case where $\tau_i = 0$ and $\theta_i = \sigma_i$ for all $i \in V$. In this instance, we only have agents of Types II and IV.

Lemma 39. *Given any seed set S, we note:*

1. $\pi(S) = \pi(S \cup T_1)$ *and*
2. $\pi(S) \leq \pi(S \cup T_3)$.

Proof. These follow from the observation that

1. Agents of Type I are those that do not look at the innovation since ϕ is below their threshold τ_i, and thus do not affect the payoff function when added to any seed set.
2. Agents of Type III are those for which $\phi \geq \theta_i$, and therefore they accept the innovation, but do not announce it to their friends. Therefore, adding such an agent to the seed set increases the payoff by exactly $p > 0$ per agent added.

□

In the simplest case, we have $\tau_i = \theta_i$ and $\sigma_i = 1$, such that agents will be either of Type I or III, and the optimal seed set is precisely $S = T_3$. That is, agents only view an innovation if they are going to accept it and they never announce it to their neighbors, and there is no cascade triggered as a result of seeding agents.

When this is not the case, we will note that the results in the previous section can be adapted to this setting to find an optimal seed set efficiently. Given a network G in this generalized setting, consider a corresponding network $G' = (V', E')$, which is the subgraph of G consisting of agents of only Types II and IV. We can then apply the algorithm in the previous section to this subgraph G' to find an optimal seed set. We claim that the union of this with agents of Type III yields an optimal seed set in G.

Theorem 40. *Given an innovation with a value ϕ and a network G with agents of parameters τ_i, θ_i, and σ_i, there is a polynomial-time algorithm for finding an optimal seed set to maximize the payoff function.*

Proof. Given such a graph G, and a corresponding subgraph G' with an optimal seed set S', we argue that $S' \cup T_3$ is an optimal seed set in G.

For the sake of contradiction, suppose S is an optimal seed set in G, such that $\pi(S) > \pi(S' \cup T_3)$. By Lemma 39, we can assume that S does not include any agent of Type I and includes all agents of Type III. This assumption implies $\pi(S \setminus T_3) > \pi(S')$. But since $S \setminus T_3 \subseteq G'$, this contradicts the optimality of S'.

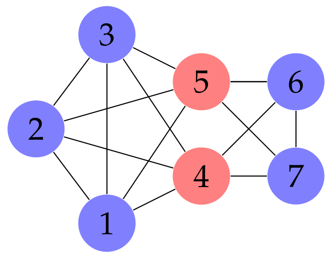
□

Returning to the implications for the overexposure phenomena, we note that a designer can efficiently maximize its payoff over the choice of *both* the seed set and ϕ. Suppose the designer knows the parameters $\tau_i, \theta_i, \sigma_i$ for each of the agents i. Given n agents, these values divide up the unit interval into at most $4n$ subintervals I_j. It is easy to see that the payoff depends only on which subinterval ϕ is contained in, but does not vary within a subinterval. Earlier, we saw the payoff need not be monotone in ϕ; but by trying values of ϕ in each of the $4n$ subintervals, the designer can determine a value of ϕ and a seed set that maximize its payoff.

8.3 Budgeted Seeding is Hard

In this section, we show that the seeding problem, even for the initial model, is NP-hard if we consider the case where there is a budget k for the size of the seed set and we want to find S subject to the constraint that $|S| \leq k$. In the traditional influence maximization literature, we leverage properties of the payoff function such as its submodularity or supermodularity to give algorithms that find optimal or near-optimal seed sets. The payoff function here, however, is neither submodular nor supermodular.

Example 41. Take the network shown in the figure below. Set $p = q = 1$



Supermodularity states that given $S' \subseteq S$ and $x \notin S$,

$$f(S' \cup \{x\}) - f(S') \leq f(S \cup \{x\}) - f(S).$$

Set $S' = \{1\}$, $S = \{1, 6\}$, and let x be node 7. Then, supermodularity would give

$$\pi(\{1, 7\}) - \pi(1) \leq \pi\{1, 6, 7\} - \pi\{1, 6\}$$
$$2 \leq 0$$

Submodularity states that

$$f(S' \cup \{x\}) - f(S') \geq f(S \cup \{x\}) - f(S).$$

A counterexample to this is obtained by setting $S' = \emptyset$, $S = \{1\}$ and $x = \{7\}$.

Here, we show an even stronger hardness result: it is NP-hard to decide if there is a (budgeted) set yielding positive payoff. Since it is NP-hard to tell whether the optimum in any instance is positive or negative, it is therefore also NP-hard to provide an approximation algorithm with any multiplicative guarantee — a sharp contrast with the multiplicative approximation guarantees available for budgeted problems in more traditional influence maximization settings.

Theorem 42. *The decision problem of whether there exists a seed set S with $|S| \leq k$ and $\pi(S) > 0$ is NP-complete.*

Proof. We will prove this using a reduction from the NP-complete Clique problem on d-regular graphs: given a d-regular graph G and a number k, the question is to determine whether there exists a *k-clique* — a set of k nodes that are all mutually adjacent. (We also require $d \geq k$.)

We will reduce an instance of k-clique on d-regular graphs to an instance of the decision version of our budgeted seed set problem as follows. Given such an instance of Clique specified by a d-regular graph G and a number k, we construct an instance of the budgeted seed set problem on a new graph G' obtained from G as follows: we replace each $(i, k) \in E$, with two new edges $(i, j), (j, k)$, where j is a new node introduced by *subdividing* (i, k). Let V be the set of nodes originally in G, and V' the set of nodes introduced by subdividing. In the seed set instance on G', we define θ_i and ϕ such that $\theta_i < \phi$ for all $i \in V$ and $\theta_i > \phi$ for all $i \in V'$. We define the payoff coefficients p, q by $q = 1$ and $p = d - (k-1)/2 + \epsilon$ for some $0 < \epsilon < 1/n^2$.

We will show that G has a k-clique if and only if there is a seed set of size at most k in G' with positive payoff.

First, suppose that S is a set of k nodes in G that are all mutually adjacent, and consider the corresponding set of nodes S in G'. As a seed set, S has k accepting nodes and $kd - \binom{k}{2}$ rejecting neighbors, since G is d-regular but the nodes on the $\binom{k}{2}$ subdivided edges are double-counted. Thus the payoff from S is $kp - (kd - \binom{k}{2}) = k(p - d + (k-1)/2)$, which is positive by our choice of p.

For the converse, suppose S has size $k' \leq k$ and has positive payoff in G'. Since the seed set consists entirely of accepting neighbors (any others can be omitted without decreasing the payoff), $S \subseteq V$, and hence so the neighbors of S reject the innovation. If S induces ℓ edges in G, then the payoff from S includes a negative term from each neighbor, with the nodes on the ℓ subdivided edges double-counted, so the payoff is $k'p - (k'd - \ell)$. If $|S| = k' < k$, then since $\ell \leq \binom{k'}{2}$, the payoff is at most $k'p - (k'd - \binom{k'}{2}) = k'(p - d + (k'-1)/2)$, which is negative by our choice of p. If $|S| = k$ and $\ell \leq \binom{k}{2} - 1$, then the payoff is at most $kp - (kd - (\binom{k}{2} - 1)) = k(p - d + (k-1)/2 - 1/k)$, which again is negative by our choice of p. Thus it must be that $|S| = k$ and $\ell = \binom{k}{2}$, so S induces a k-clique as required. □

8.4 Experimental Results

In this section, we present some computational results using data sets obtained from SNAP (Stanford Network Analysis Project). In particular, we consider an email network from a European research institution (Leskovec et al., 2007b; Paranjape et al., 2017) and a text message network from a social-networking platform at UC-Irvine (Panzarasa et al., 2009).

The former is a directed network of emails sent between employees over an 803-day period, with 986 nodes and 24929 directed edges. The latter is a directed network of text messages sent between students through an online social network at UC Irvine over a 193-day period, with 1899 nodes and 20296 directed edges. In both networks, we use the edge (i, j) to indicate that i sent at least one email or text to node j over the time period considered.

For both of these networks, we present results corresponding to the general model. We consider 100 evenly-spaced values of ϕ in $[0, 1]$ and compare the seed set obtained by our algorithm with some natural baselines. The parameters $\tau_i \leq \theta_i \leq \sigma_i$ for each agent are chosen as follows: we draw three numbers independently from an underlying distribution (we analyze both the uniform distribution on $[0, 1]$ and the Gaussian distribution with mean 0.5 and standard deviation 0.1); we then sort these three numbers in non-decreasing order and set them to be τ_i, θ_i, and σ_i respectively.

For each ϕ we run 100 trials and present the average payoff. The average time to run one simulation is 0.915 seconds for the text network and 0.454 seconds for the email network. This includes the time to read the data and assemble the network; the average time spent only on computing the min-cut for the corresponding network is 0.052 and 0.054 seconds respectively.

For each of these figures, we give a natural upper-bound which is the number of agents such that $\theta_i < \phi$. This includes agents of Type III and IV. We note that the seed set obtained by our algorithm often gives a payoff close to this upper bound. We compare this to two natural baselines: the first sets agents of Type III to be the seed set and the second sets agents of both Types III and IV to be the seed set. We show that the first baseline performs well for lower values of ϕ, where the second baseline underperforms significantly; and, the second picks up performance significantly for higher values of ϕ while the first baseline suffers. The seed set obtained by our algorithm, on the other hand, outperforms both baselines by a notable margin for moderate values of ϕ. This gap in performance corresponds to the overexposure effect in our models.

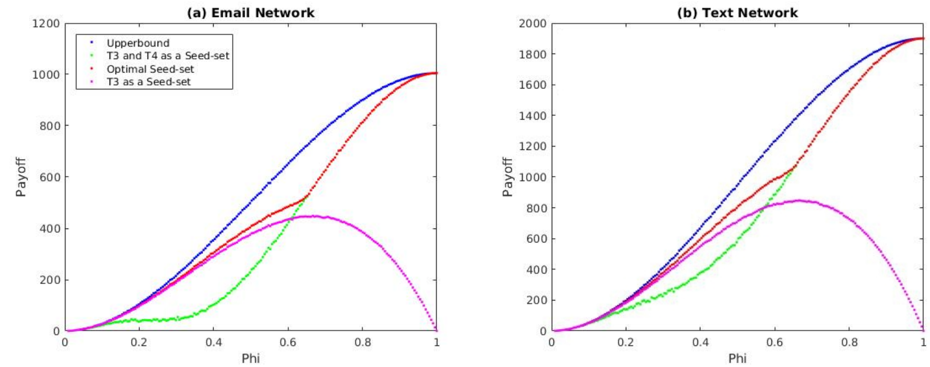


Figure 8.1: Payoff as a function of the quality ϕ for (a) email and (b) text networks for the general model where $\tau_i, \theta_I, \sigma_i$ are chosen from the uniform distribution on $[0, 1]$. The blue curve represents the natural upper bound of the total number of agents with $\theta_i < \phi$. The green line corresponds to the payoff obtained by seeding agents whose $\tau_i \leq \phi$. The purple line represents the number of agents of Type III, where $\phi \in [\theta_i, \sigma_i)$ and the red line corresponds to the seed set selected by our algorithm. The difference between the red and the green curve capture the overexposure phenomena.

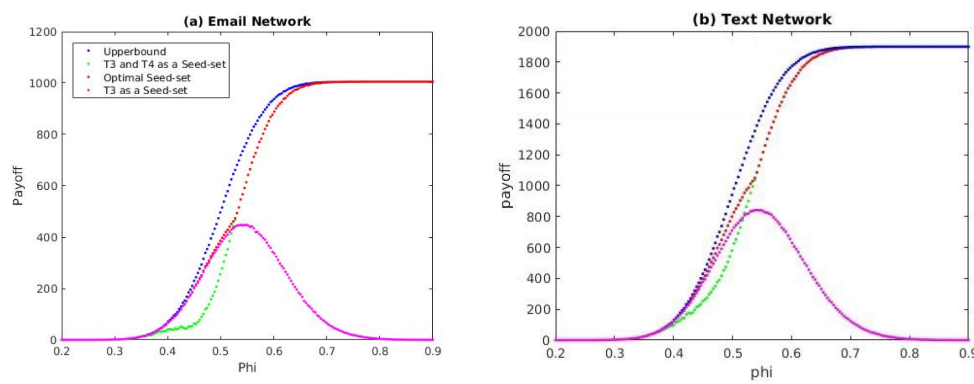


Figure 8.2: These plots give payoff for (a) email and (b) text networks for the general model with parameters chosen from a Gaussian distribution with mean 0.5 and standard deviation 0.1 for 100 evenly spaced ϕ values in $[0, 1]$.

Each of the figures show that the seed set chosen by our algorithm outperforms these natural baselines. In Figure 8.1, we note for $\phi < 2/3$ the optimal seed set obtained through our algorithm is close to picking only agents of Type III. Adding on agents of Type IV performs worse than both seeding just agents of Type III or the optimal seed set. This changes for ϕ values over 2/3. Here, the number of agents of Type III drops, and thus the payoff obtained by seeding agents of Type III drops with it. On the other hand, the seed set consisting of all agents of Types III and IV picks up performance, coming close to the optimal seed set for $\phi \approx 0.7$. This behavior appears in both networks.

8.5 Discussion

Theoretical models of word-of-mouth recommendations in social networks have generally used the assumption that all exposures to an innovation are beneficial to the designer. A separate line of empirical research in public health campaigns and marketing, however, provide a more complex picture, in which different potential individuals may have either positive or negative reactions to an innovation, and it can be a mistake to pursue a strategy that elicits too many negative reactions from the population.

In this chapter, we have proposed a new set of theoretical models for word-of-mouth recommendations, by taking into account these types of overexposure effects. Our models make it possible to consider the optimization trade-offs that arise from trying to reach a large set of positively inclined potential individuals while reducing the number of negatively inclined potential individuals who are reached in the process. Even in the case where the designer has no budget on the number of people it can expose to the innovation, this tension between positive and negative reactions leads to a non-trivial optimization problem. We provide a polynomial-time algorithm for this problem, using techniques from network flow, and we prove hardness for the case in which a budget constraint is added to the problem formulation. Computational experiments show how our polynomial-time algorithm yields strong results on network data.

Our framework suggests many directions for future work. It would be interesting to integrate the role of negative payoffs in our model here with other technical components that are familiar from the literature on influence maximization, particularly the use of richer (and potentially probabilistic) functions governing the spread from one participant in the network to another. For example, when nodes have non-trivial thresholds for adoption — requiring both that they evaluate the innovation positively and also that they have heard about it from at least k other people, for some $k > 1$ — how significantly do the structures of optimal solutions change?

It will also be interesting to develop richer formalisms for the process by which positive and negative reactions arise when potential individuals are exposed to good or bad innovations. With such extended formalisms we can more fully bring together considerations of overexposure and reputational costs into the literature on network-based word-of-mouth recommendations.

CHAPTER 9

FACILITATING NETWORK INTEGRATION

Across societies, the tendency for individuals to form social ties with others with whom they share similarities is a robust and pervasive social phenomena impacting the formation of social networks (Lazarsfeld et al., 1954; Kossinets and Watts, 2009; McPherson et al., 2001; Newman, 2002). This phenomena, known as *homophily*, can result in segregation; and since networks play a key role in the diffusion of information, opportunities, and resources, it can lead to inequalities across members of different communities; empirical and theoretical work have shown that network structures can influence individuals' ability to obtain accurate and relevant information, garner social support, improve their labor market outcomes, among other impacts (Calvo-Armengol and Jackson, 2004; Calvo-Armengol et al., 2009; Jackson et al., 2012; Banerjee et al., 2013; Stoica et al., 2018; Nilizadeh et al., 2016; Hannák et al., 2017).

Focusing on access to information related to jobs, for instance, a long line of work in economics and sociology has looked at both theoretical models and empirical evidence showing what impact networks can have on attaining jobs (Calvo-Armengol and Jackson, 2004; Fernandez and Fernandez-Mateo, 2006; Pedulla and Pager, 2017). Homophily and resulting segregation on social networks can impact diffusion of information on social networks leading to disparities in access to employment opportunities. And, more generally, social network and social ties have been used as a means of attaining status more generally leading to improvement in access to opportunities (Grusky, 2018; Lin, 1999).

Understanding how network formation processes are impacted by homophily is important for predicting and improving societal welfare in various domains. However, since homophily is a potent and organic force, it is challenging to push back against its negative consequences without identifying other social processes that may already be working against segregation. By pinpointing such network forces and understanding their interactions with homophily, we can then harness their impact to improve network diversity and welfare of individuals across the network.

In this chapter, we show that *triadic closure* – the process by which individuals are more likely to form social ties with other individuals with whom they already share ties – may be one such force (Granovetter, 1977; Kossinets and Watts, 2006; Rapoport, 1953). Through analyses of different popular network formation models, we show that triadic closure has the tendency to reduce network segregation. That is, even though biases towards homophily in link formation tend to push a network towards segregation, triadic closure has the effect of exposing people to dissimilar individuals, thereby resulting in more integrated networks. Indeed much work, including recent work, has asserted that triadic closure can result in social segregation at the macroscale (Stadtfeld, 2015; Tóth et al., 2019).

We find it striking that such a tension should exist between two such well-studied social processes as homophily and triadic closure, and we believe it points to the possibility of broader phenomena that we begin exploring in this chapter. The tension between these two forces is also partly counter-intuitive: triadic closure produces links between people with common friends, and we might therefore suspect that these new links reinforce the underlying similarity that the friends share; but the longer-range nature of triadic closure, connecting people separated by multiple steps in the network, in fact has the aggregate effect of exposing nodes to others who are dissimilar.

We explore this relationship between these phenomena in two ways: we first analyze the effect of triadic closure on the proportion of links between dissimilar agents on different static and dynamic network formation models. Specifically, we focus on a static network formation model based on stochastic block models (SBM) and a dynamic model generalizing a popular model by Jackson and Rogers (2007). Both models are characterized by the presence of heterogeneous nodes and a two phase link-formation process where agents first form a fixed set of links followed by a triadic closure phase were they form an additional set of links through a friend-of-friend search. In both cases, we find that triadic closure increases network integration if agents display homophily in the first phase.

We then explore network interventions that leverage these insights. Namely, in many scenarios, an authority that has a vested interest in the outcome of a network process can attempt to nudge the network towards a more integrated state. For example, many universities randomize dormitory assignments or the formation of freshman groups at the beginning of a year in order to encourage a diversity of links. On-line networking platforms recommend links to create a warm start for new users, as well as to fill out established networks. In most contexts, one's power to control the global network structure is quite limited; while the authority might want the network to be highly integrated, it is unlikely that they can influence the structure of all relationships on a global scale. Here, we show that the above insights may help overcome these limitations; minor and local interventions on homophily in the first phase can be amplified through triadic closure, resulting in an indirect yet substantial shift on the network integration.

The remainder of the section is organized as follows. We begin with an analysis of triadic closure in the stochastic block model, deriving mathematical results on its effect on network integration. We then introduce and analyze a growing graph model based on the Jackson-Rogers model, considering the effect of triadic closure on network integration in an equilibrium state. We close with a discussion on the design of interventions that act on the initial phase to optimize long-term network integration. We support these results with experiments and discussions about the interplay of homophily, triadic closure, integration, and implications for network interventions on- and off-line settings.

9.1 Triadic Closure and Stochastic Block Models

We begin with a simple static model based on stochastic block models (SBM). We suppose that there are n nodes corresponding to individuals, where each node can be one of k types. We assume that there are an equal number of nodes of each type. For ease of presentation, we assume that $k = 2$ where the two types are $\{R, B\}$. We present expanded proofs for $k > 2$ in the appendix.

In an SBM with $k = 2$ types, there is a probability p of an edge forming between two nodes of the same type (this is the *in-block probability*) and a probability q of an edge forming between two nodes of different types (this is the *out-of-block probability*). All edge formations are mutually independent. We first run an SBM to obtain a network G. We then consider the effect of triadic closure on G. In order to do so, we must study wedges in G.

A *wedge* W in a graph G is an induced subgraph on nodes $\{v_1, v_2, v_3\}$ such that there exists edges (v_1, v_2) and (v_2, v_3), but there exists no edge between nodes v_1 and v_3. We denote this wedge W using (v_1, v_2, v_3).

Definition 43. We say $W = (v_1, v_2, v_3)$ is a *monochromatic wedge* if v_1 and v_3 share a type. Otherwise, we say it is a *bichromatic wedge*.

Definition 44. Given a set of nodes $T = \{v_1, v_2, v_3\}$, we say that T is a *monochromatic triplet* if all three nodes share a type and we say that it is a *bichromatic triplet* otherwise.

Monochromatic and bichromatic wedges correspond to wedges where the missing edges are monochromatic and bichromatic, respectively. Note, further, that a monochromatic wedge can result from either a monochromatic or a bichromatic triplet, whereas a monochromatic triplet can only yield a monochromatic wedge. An object of interest in this section is the fraction of bichromatic wedges, which we denote by $w(G)$. We can estimate this value using the following result:

Lemma 45. *Given a network G resulting from an SBM, the expected number of monochromatic wedges is*

$$3 \cdot 2 \cdot \binom{n/2}{3} \cdot p^2(1-p) + n \cdot \binom{n/2}{2} \cdot q^2(1-p), \tag{9.1}$$

while that of bichromatic wedges is

$$2 \cdot n \cdot \binom{n/2}{2} \cdot pq(1-q). \tag{9.2}$$

Proof. The expected number of monochromatic wedges from monochromatic triplets is

$$3 \cdot 2 \cdot \binom{n/2}{3} \cdot p^2(1-p),$$

where $\binom{n/2}{3}$ counts the number of monochromatic triplets of each type, 3 is the number of ways to choose the center of the wedge, and $p^2(1-p)$ counts the likelihood that such a triplet results in a wedge. The extra factor of two reflects the $k = 2$ blocks.

Similarly, the expected number of monochromatic wedges from a bichromatic triplet is

$$n \cdot \binom{n/2}{2} \cdot q^2(1-p),$$

where $2(n/2)\binom{n/2}{2}$ counts the number of bichromatic triplets and $q^2(1-p)$ corresponds to the likelihood that such a triplet results in a monochromatic wedge, thereby giving us the desired result in Equation 9.1.

As for the expected number of bichromatic wedges, we have already noted that there are $n\binom{n/2}{2}$ bichromatic triplets. Such a triplet results in a bichromatic wedge with probability $2pq(1-q)$, again yielding the desired result in Equation 9.2. □

We consider the effect of triadic closure on network integration. We measure network integration using the proportion of bichromatic edges: i.e., the fraction of edges in G between nodes of dissimilar types, denoted by $f(G)$.

Definition 46. We say that triadic closure *increases network integration* if, by closing a single randomly selected wedge, we increase the fraction of bichromatic edges. We say that triadic closure *decreases network integration* if this action decreases the fraction of bichromatic edges. And we say that triadic closure *preserves network integration* if this action does not change the fraction of bichromatic edges.

This above definition is a local one: it analyzes the impact of a single triadic closure. Note, however, that for large n, this observation on network integration applies for the case where we can instead consider the effect of an unbounded but sublinear number of simultaneous triadic closures.

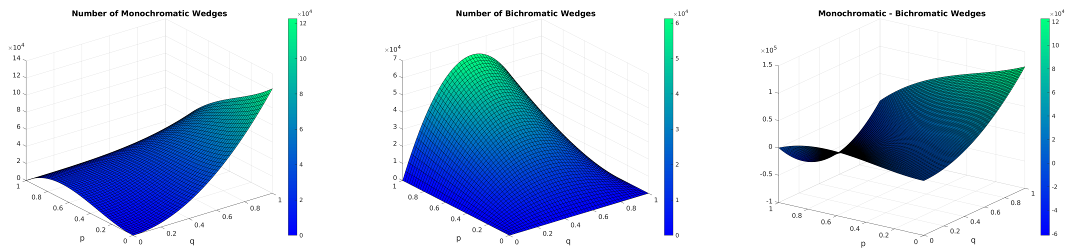


Figure 9.1: Surface plots showing the number of monochromatic and bichromatic edges as we vary p and q in an SBM.

Theorem 47. *Given an SBM, there is a sufficiently large n such that triadic closure increases network integration if $p > q$ and it preserves network integration if $p = q$.*

Proof. We are interested in both the fraction of bichromatic edges, $f(G)$, and the fraction of bichromatic wedges, $w(G)$. First, note that

$$f(G) = \frac{(n/2)^2 q}{2 \cdot \binom{n/2}{2} \cdot p + (n/2)^2 \cdot q} = \frac{q}{p+q} + o(1),$$

where the numerator is the expected number of bichromatic edges and the denominator is the expected number of all edges.

On the other hand, using the previous lemma and simplifying the expression, we note that the fraction of bichromatic wedges is

$$w(G) = \frac{pq(1-q)}{pq(1-q) + \frac{p^2(1-p)}{2} + \frac{q^2(1-p)}{2}} + o(1).$$

Note that if triadic closure is applied to a random wedge, the probability that the resulting edge is bichromatic is $w(G)$. This increases the fraction of bichromatic edges if $f(G) < w(G)$, thus increasing network integration, and similarly it preserves network integration if $f(G) = w(G)$.

We therefore study the inequality $f(G) \leq w(G)$ for sufficiently large n. We have

$$\begin{aligned}
\frac{q}{p+q} &\leq \frac{pq(1-q)}{pq(1-q) + \frac{p^2(1-p)}{2} + \frac{q^2(1-p)}{2}} \\
\frac{qp^2(1-p) + q^3(1-p)}{2} &\leq p^2q(1-q) \\
p^2(1-p) + q^2(1-p) &\leq 2p^2(1-q) \\
q^2(1-p) &\leq p^2(1-2q+p)
\end{aligned}$$

Note that this last line holds with strict inequality if $p > q$ and with equality if $p = q$; triadic closure increases network integration if nodes display homophily and it preserves network integration when there is no homophily. □

We note the following necessary subtlety in the proof: an SBM with $p > q$ always has $f(G) < 1/2$ for sufficiently large n, since $f(G) = q/(p+q) + o(1)$. Thus, one might imagine a strategy for proving this theorem that sought to show $w(G) \geq 1/2$, which would be sufficient. However, as we can see in the Figure 9.1, this is not always the case – in particular, for low values of q and high values of p, there can be more monochromatic wedges than bichromatic wedges, and hence $w(G) < 1/2$. Given this observation, we must therefore focus on a more careful analysis of the relative sizes of $f(G)$ and $w(G)$ to obtain the above result.

9.2 A Dynamic Network Formation Model

We now consider a dynamic process of network formation, based on a natural variation of the popular growing graph model by Jackson and Rogers (2007).

In the Jackson-Rogers model, homogeneous nodes arrive sequentially and form links first through a *random meeting phase* and next through a *network search phase*.[1] We adapt this model to account for heterogeneous nodes. As in the previous section, we assume that each node can be one of two types and that there are an equal number of nodes of each type.

The model proceeds as follows: nodes arrive to the network consecutively. Upon arrival, each node v forms undirected links in two phases: in the first phase, it meets a fixed number of agents of each type. In the second phase, the node meets friends-of-friends to create further links, which we can view as an organic development of new connections through one's friend. We denote the resulting network after the friendship formation of the t'th arriving node by $G(t)$. More specifically, when node v arrives at time t,

I it selects N_S neighbors uniformly at random from nodes of the same type and N_D neighbors uniformly at random from nodes of a different type, then

II it selects N_F additional neighbors according to the following biased friend-of-friend search process: suppose $\alpha \in (0, 1]$. Let $F_S(v)$ be the set of same-

[1]The original Jackson-Rogers model differentiates between meeting and link formation in the first phase, where links result from meeting based on compatibility. Consistent with a lot of follow up work on this model, we do not differential between meeting and link formation and assume that random meeting results in link formation. Note that we do not lose generality in doing so since we can easily convert the first phase in this model to the original Jackson-Rogers model.

type neighbors that v selected in the first phase and $F_S^2(v) = \{w : u \in F_S(v),\ (u, w) \in G(t-1)\}$ be the neighbors of v's same-type neighbors. (We can similarly define $F_D(v)$ and $F_D^2(v)$.) Then v selects αN_F nodes uniformly at random among $F_S^2(v)$ and $(1-\alpha)N_F$ nodes uniformly at random among $F_D^2(v)$ and creates links.

We let $N := N_S + N_D + N_F$ and so the number of edges in $G(t)$, which we denote by $m(t)$, is Nt. We assume that N_S, N_D, N_F are all fixed constant values. Our results hold for arbitrary values of N_S, N_D, and N_F, but we will be especially interested in the regime where the number of links formed in the random meeting phase is small compared to the search phase, motivated by the discussion below.

We can view the random meeting phase as an initial "seeding" phase of the individual's entry into the community, and one should imagine that without intervention, there will be a bias toward connecting to more individuals of a similar type. This models a direct bias toward homophily. Mirroring real-world network formation processes, we can think of the random meeting phase as dorm assignments in a university dorm or links recommended to a new user when they arrive on an online platform.

The search phase, as argued in Jackson and Rogers (2007), is a natural abstraction of link formation via search: individuals are naturally exposed to friends of their friends, and some of these meetings lead to the formation of connections via triadic closure. This mirrors formation of friendships on a college campus, link formation on social media, and how citation and scientific collaboration networks form and evolve. We model a homophily bias as being present in this search procedure by assuming that individuals are more likely to connect with friends of same-type friends. i.e., we can assume that $\alpha > \frac{N_S}{N_S+N_D}$, although the results below hold for any $\alpha > 0$. This assumption captures an indirect form of bias, where one tends to disproportionately trust the friend recommendations from same-type individuals. One might expect such a process to amplify any homophily due to the bias in the first phase, as an abundance of same-type links will result in any accepted friend-of-friend connections also having a higher chance of being same-type.

In what follows, we show that triadic closure again has the effect of *increasing network integration* if nodes exhibit homophily in the first phase. We first show that as t goes to infinity, the fraction of bichromatic edges converges to an equilibrium state. We analyze the effect of both the random meeting and network search phase on the rate of convergence and further show that search phase characterized by triadic closure mitigates network segregation.

9.2.1 Network Integration at Equilibrium

We denote the equilibrium state network, that is $G(t)$ as $t \to \infty$, by G. To ensure that this process is well defined, we let $G(0)$ a stochastic block model on N nodes and we let each node select $\frac{N_F}{N_D+N_S}$ nodes to form links with uniformly at random. We assign each node a random type.

We are interested in the network integration, as measured by the fraction of bichromatic edges, in equilibrium. We denote by $f(t)$ the fraction of bichromatic edges at time t. Here, we want to show that $f(t)$ converges and that we can fully characterize its value in equilibrium.

Fraction of Bichromatic Edges

Let $b(t)$ be the number of bichromatic edges at time t and recall that $m(t)$ is the number of all edges. Therefore, $f(t) = b(t)/m(t)$.

Given a node v, we let $f_v(t)$ be the fraction of v's neighbors that do not share a type with v. We can then define

$$f_R(t) = \frac{1}{|R|} \sum_{r \in R} f_r(t).$$

Note that this is an average value over all nodes of type R. We can similarly define $f_B(t)$.

Since this is a growing graph, the degree of nodes also evolves. We denote the degree of node v at time t by $\Delta_v(t)$ and define

$$\Delta_R(t) = \frac{1}{|R|} \sum_{r \in R} \Delta_r(t).$$

That is, $\Delta_R(t)$ is the average degree of an R node at time t. Therefore,

$$b(t) = \sum_{r \in R} \Delta_r(t) f_r(t) = \sum_{b \in B} \Delta_b(t) f_b(t).$$

Without loss of generality, suppose a new node v of type R is added at time $t+1$. Then,

$$\begin{aligned} b(t+1) = b(t) &+ N_D + \alpha N_F f_R(t) \\ &+ (1-\alpha) N_F (1 - f_B(t)) \end{aligned} \tag{9.3}$$

Note that the first term counts the number of bichromatic edges already present at time t, N_D counts the number of new bichromatic edges introduced in the random meeting phase, $\alpha N_F f_R(t)$ counts the expected number of new bichromatic links that v forms via triadic closure through same-type neighbors, and

$(1-\alpha)N_F(1-f_B(t))$counts the expected number of new bichromatic links that v forms via triadic closure through their neighbors of a different type. On the other hand, the number of all edges goes up by N.

$$m(t+1) = m(t) + N.$$

Our analysis involves a step in which we establish an identity (Identity 1, below) on the long-run fraction of bichromatic edges $f(t)$ using a heuristic argument that is supported by numerical simulation.

The identity is as follows. Let f^* denote the limiting fraction of bichromatic edges of $G(t)$ as t goes to infinity. The we have

$$\text{(Identity 1)} \qquad f^* = \frac{N_D + (1-\alpha)N_F}{N_D + N_S + 2(1-\alpha)N_F}.$$

We note that $f(t)$ shows that the fraction of bichromatic links formed in the first phase can have a disproportionate impact on the network integration in equilibrium. For instance, for $\alpha = 1$, which corresponds to the instance where all friend-of-friend links are formed through same-type friends, $f(t) = \frac{N_D}{N_S+N_D}$, indicating that this value is *exclusively* determined by the first phase despite the assumption that $N_S + N_D$ is much less than N_F. The role of N_F in the fraction of bichromatic edges in equilibrium increases as α decreases. Further, for *any* $\alpha < 1$, $f(t)$ approaches $1/2$ as N_F grows large relative to N_S and N_D. So even a single off-type seed friend can have a significant down-stream impact, if triadic closure is not perfectly driven by same-type friends. We will leverage these insights to explore simple but robust interventions to increase network integration in the following section.

Triadic Closure and Network Integration

Mirroring the discussion in the previous section, we say that triadic closure *increases network integration* if f^* goes up as N_F goes up, we say that it *decreases network integration* if f^* goes down as N_F goes down, and that it *preserves network integration* if f^* is unaffected by N_F.

Recall that we assume that nodes show type-bias: i.e., $N_S < N_D$ and $\alpha > \frac{N_S}{N_D+N_S}$, corresponding to type-bias in the random meeting and network search phases, respectively.

Theorem 48. *Assuming Identity 1, triadic closure increases network integration in our dynamic network formation model if $\alpha < 1$ if nodes show type-bias in the first phase and preserves network integration if nodes show no bias. Triadic closure preserves network integration for $\alpha = 1$.*

Proof. Select $N_F, N_{F'}$ such that $N_F < N_{F'}$. We aim to understand the effect of forming N_F versus $N_{F'}$ links in the second phase on the respective equilibrium values for the fraction of bichromatic edges, which we denote by f^* and f'^*, respectively. Specifically, we are interested in the inequality $f^* \leq f'^*$, which we can rewrite as

$$\frac{N_D + (1-\alpha)N_F}{N_D + N_S + 2(1-\alpha)N_F} \leq \frac{N_D + (1-\alpha)N_{F'}}{N_D + N_S + 2(1-\alpha)N_{F'}}$$

We can simplify this above inequality to get

$$2N_D(N_{F'} - N_F) \leq (N_S + N_D)(N_{F'} - N_F).$$

This holds with strict inequality if $N_D < N_S$ and with equality if $N_D = N_S$, giving us the desired result for $\alpha < 1$.

For the degenerate case where $\alpha = 1$, note f^*, f'^* are both $\frac{N_D}{N_S+N_D}$. That is, the fraction of bichromatic edges is determined exclusively by the random meeting phase and triadic closure preserves network integration in this setting. $\square$

Note that for the case where $N_S = N_D$, triadic closure preserves network integration even if nodes show type-bias in the second phase, by forming a disproportionate number of links through their same-type friends. Likewise, for the case where $\alpha = 1$, triadic closure preserves network integration regardless of the type-bias that nodes show in the first phase. Outside of these two degenerate cases, we note that triadic closure has the effect of increasing network integration.

Rate of Convergence

Finally, we find that we can also explicitly state how each of these parameters impact the rate at which the network converges to the equilibrium state.

At equilibrium, $f(t) = f(t+1)$. We are therefore interested in

$$\begin{aligned} & f(t+1) - f(t) \\ & \quad = \frac{N_D + \alpha N_F f(t) + (1-\alpha) n_F (1 - f(t))}{N_S + N_D + N_F} - f(t) \end{aligned}$$

We solve the differential equation,

$$\frac{df(t+1)}{dt} = \frac{N_D + 2\alpha N_F f(t) + n_F(1 - f(t))}{N_S + N_D + N_F} - f(t)$$

to get that,

$$f(t) = \frac{N_D + N_F - c \cdot \exp(-(t(N_F - 2\alpha N_F))/(N)}{N_S + N_D + 2(1-\alpha)N_F}$$

where c is some constant. In addition to impacting the equilibrium state, triadic closure also affects the rate of convergence to this state. This holds even for the case where $\alpha = 1$ when triadic closure preserves network integration.

9.3 Interventions to Alleviate Network Segregation

Informed by the observations on the impact of triadic closure on network integration, we explore potential interventions aimed at decreasing segregation. We study interventions on the dynamic model from the previous section. In this model, we have noted that links formed in the first phase can have an out-sized effect on the eventual equilibrium state of the network even if we assume that they comprise a small fraction of the total number of edges formed.

Here, we focus on interventions that act solely on the first phase. Recalling our motivating examples related to college dormitory assignments or recommendation of friendships when an individual joins an online platform, we note that an authority (i.e., university or platform, respectively) may have more leverage in this initial phase than subsequent steps which proceed through friend-of-friend searches. Such interventions that act as "nudges" in the initial phase have recently been popular in the fairness in recommender systems community; research in this space has explored the impact of bias in link formation or other selection on the long-term health of online platforms, with some work exploring the role of small nudges by the platform to mitigate inequalities or achieve other desirable social outcomes (Ekstrand and Willemsen, 2016; Guy, 2015; Hutson et al., 2018; Knijnenburg et al., 2016; Schnabel et al., 2018; Stoica et al., 2018; Su et al., 2016).

In our analysis of interventions, we assume that the number of links formed in the first phase is fixed but that the designer has the ability to change the proportion of mono- versus bi-chromatic edges formed in the initial seeding phase subject to this sum constraint.[2] This intervention imitates, for instance, dorm assignments where there is a fixed number of slots per dorm but universities have the ability to change the composition of occupants in each dorm. We may also consider the setting where the designer would like to optimize network integration subject to rate-of-change constraints on the network or on the time-frame over which the intervention can occur. This constraint models scenarios where it may be costly, infeasible, or undesirable to introduce a dramatic change all at once.

Here, we show that we can find simple mechanisms for optimizing network integration building on insights from the previous section. The interventions support the evidence from the fairness in recommender systems literature that small interventions upon a user's arrival to an online platform can have long-lasting impact, even in cases where the user may engage in substantially more

[2]Note that it is straightforward to generalize the results in this section to the setting where there are no constraints on the sum $N_S + N_D$, but we discuss the case where it is for ease of presentation.

activity after this initial contact.

We imagine here that the designer has an ideal level of integration that they would like to see in equilibrium, which we denote by $\hat{f}$. The designer would like to reach this state by acting on only the first phase. The optimal intervention in this setting is an immediate consequence of the previous lemma:

Corollary 49. Given a network $G(0)$, a budget $N - N_F$ for links in the first phase, and a desired level of integration $\hat{f}$, the intervention which solves for N_D in the following equation

$$\hat{f} = \frac{N_D + (1-\alpha)N_F}{N - N_F + 2(1-\alpha)N_F}.$$

yields an integration level of $\hat{f}$ in the equilibrium state.

This observation will serve as a building block for finding solutions that take into account to natural considerations: rate-of-change of the network structure and time-frame over which the intervention can occur.

We first consider the case where the objective is to maximize the network integration at time t. Before presenting this result, we note the following observation: given a network $G(t)$ with fraction of bichromatic edges $f(t)$, the intervention which maximizes the fraction of bichromatic edges at time $t+1$ is that which sets $N_D = N - N_F$. This intervention maximizes integration *locally*. We also note that this intervention maximizes integration over a longer time-horizon.

Lemma 50. *Suppose we are given a network $G(0)$, values N_F, N, and α, and a time-horizon $[t_1, t_2]$. Starting from $f(t_1)$, the intervention which maximizes $f(t_2)$ is that which sets $N_D = N - N_F$ over the entire time-horizon $[t_1, t_2]$.*

Central to the argument for this above result is Figure 9.2, which illustrates the feasible region for the fraction of bichromatic edges. A point on the graph corresponds to a tuple $(t, f(t))$. The green lines correspond to the interventions which set $N_D = N - N_F$ and the red lines correspond to those which set $N_D = 0$. Given $G(t_1)$, the set of possible values $f(t)$ for $t \in [t_1, t_2]$ can be traced using a path from $(t_1, f(t_1))$ to (t_2, x) where x must be in the feasible region and the path must follow the grid taking no left steps.

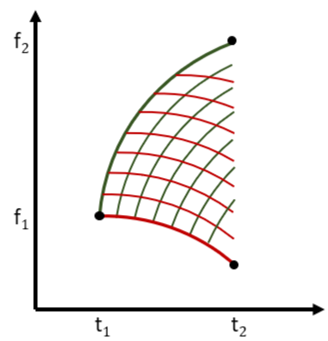


Figure 9.2: Plot showing the feasible region of fraction of bichromatic edges in our network during time period $[t_1, t_2]$.

While the above intervention allows us to reach the optimum integration level given a fixed time-horizon, it may be costly to implement: namely, the designer may wish to introduce interventions gradually rather than all at once. This suggests a natural generalization of the above problem, stated below:

Theorem 51. *Given a network $G(t_1)$, values N_F, N, and α, and a time-horizon $[t_1, t_2]$, suppose the designer aims to reach an integration level f_2 by time t_2 subject to a rate-of-change constraint Δ. We can find an optimal intervention acting on phase 1 or prove that no such intervention exists.*

Proof. We first draw the feasible region as in Figure 9.2. It is a straightforward consequence of the above lemma that if (t_2, f_2) is not in the feasible region, then no such intervention exists. Suppose that (t_2, f_2) is in the feasible region. We denote by $\mathcal{L}(t_1, f_1, t_2, f_2)$ the set of all paths $\{\ell_1, \ell_2, \dots \ell_h\}$, which from (t_1, f_1) to

(t_2, f_2) that do not take any left steps. For each path, we check whether there exists any $t_i \in (t_1, t_2]$ such that $|f(t_i) - f(t_{i-1})| > \Delta$ and remove all paths that violate this rate-of-change constraint. The remaining paths each correspond to feasible interventions that abide by the given constraints. $\square$

These interventions directly take insights from the study on the interplay of homophily and triadic closure to propose simple mechanisms that can improve long-term health of a network. The simplicity of these mechanisms, coupled with both theoretical evidence in this work and empirical evidence from the recommender systems literature, suggests new directions at the interface of network science and recommender systems, which we discuss in the next sections.

9.4 Discussion

In this chapter, we consider the effect of triadic closure on network segregation. Through analysis of different static and dynamic network formation models, we find that triadic closure has the effect of increasing network integration, indicating that it may be a process that counteracts homophily in network formation.

We view the insights in this chapter as potentially pointing to broader phenomena about these social processes, which can be studied both mathematically as well as analysis of real-world networks. The results presented in this chapter open up questions related to other measurements of network health, such as network expansion and distribution of network centralities. Each of these points to challenging analytic questions. Empirically, it would also be interesting to shed light on what types of social and information networks tend to exhibit a stronger relationship between triadic closure and homophily.

The interventions presented in this chapter point to a broader set of theoretical and empirical questions. For instance, it would be interesting to estimate the various network parameters and compare the effect of nudges across different distributions of values. Furthermore, such interventions are often costly to the designer or may incur social cost, leading to a set of optimization questions where the designer must trade off these costs with utility gained from network integration. We can empirically test the impact of triadic closure on network integration on a set of social and information networks. In Section 6, we present a set of further questions on the impact of networks to increase or decrease socioeconomic inequalities inspired by the results in this section as well as the economics and sociological work from which this work transpired.

Part IV

Roles for Computing in Social Change

CHAPTER 10

OVERVIEW OF PART IV

In high-stakes decision-making, algorithmic systems have the potential to predict outcomes more accurately and to allocate scarce societal resources more efficiently. But the introduction of algorithms into these domains has also led to concerns about their potential to introduce, perpetuate, and worsen inequality. Algorithmic accountability, fairness, and bias have quickly become watchwords of mainstream technical research and practice, gaining currency both in computer science and in technology companies. And these concerns are having a public moment, with policymakers and practitioners newly attuned to their import.

Scholars have long worked to draw attention to the values necessarily embodied in technology. This call for recognition of values has been a core project of Values-in-Design (Friedman and Hendry, 2019; Introna and Nissenbaum, 2000), legal scholarship (Citron, 2018; Lessig, 2009; Reidenberg, 1997), science and technology studies (Agre, 1997; Jasanoff, 2006; Winner, 1980), and allied disciplines for decades — and remains an extremely active area of research. (For a partial listing of work drawing on such themes, see Gillespie and Seaver.) Computer science has recently been increasingly heeding the call for attention to values, politics, and "social good" more generally — with more than twenty technical workshops, conferences, or other significant gatherings on some variation of fairness, bias, discrimination, accountability, and transparency in the last five years (e.g., (Abebe and Goldner, 2018a; De-Arteaga et al., 2018; Ekstrand and Levy; Hager et al., 2019; Gomes et al., 2019; Rolnick et al., 2019)).

The proliferation of scholarly work on these questions has sparked debate about the relationship between computing and social change — in particular, about the degree to which technical interventions can address fundamental problems of justice and equity. Scholars have recently raised concerns about taking a computational lens to certain social problems, questioning whether modifications to automated decision-making can ever address the structural conditions that relegate certain social groups to the margins of society (Broussard, 2018; D'Ignazio and Klein, 2019; Crawford; Eubanks, 2018a; Gebru, 2019; Green, 2018). On these accounts, realizing principles of equity and justice requires addressing the root social, economic, and political origins of these problems — not optimizing around them (Barabas et al., 2017; Costanza-Chock, 2018; Whittaker et al., 2018; Pasquale, 2018; Powles, 2018; Selbst et al., 2019; Overdorf et al., 2018; Greene et al., 2019; Dencik et al., 2018; Hoffmann, 2019; Peña Gangadharan and Niklas, 2019; Bennett and Keyes, 2019). For instance, rather than a computational intervention that aims to equalize offers of college admission across demographic groups, we might more ambitiously direct attention to the unmet need for high-quality high school instruction in low-income neighborhoods. Similarly, an intervention at the selection phase in an employment context might mask the impact of a hostile workplace culture or other barriers to long-term employee success.

The community that has coalesced around this research interface increasingly recognizes these tensions, and is beginning to grapple with the complex relationship between computing research inspired by social problems and meaningful progress on those problems. To date, however, the field has not mapped the variety of ways in which its own work can participate in making a practical and positive difference toward social change.

The upcoming chapter presents a start to that effort: we describe four different modalities through which, alongside other efforts that support social change, computing research can play a specific and useful role in addressing social problems. These are approaches that can empower other agents of social change.

CHAPTER 11

ALTERNATIVE ROLES FOR COMPUTING

Meaningful advancements toward social change are the work of many hands. We highlight these possibilities while also considering, and taking seriously, recent critiques of computing research that attempt to address social problems. In so doing, we seek to complement the long history of bringing the critical perspectives of other disciplines to bear on computing. This work suggests that computing *itself* can be marshaled to cast a critical eye on social problems.

There is a long-standing tension between strategies that seek to intervene within the contours of an existing system, and those that seek more wholesale social and political reform. Existing work has made clear how computational approaches may contribute to the former style of activity. Here, we ask whether, and to what extent, computing can contribute to the latter activity as well. We pose this question while recognizing the critical scholarship we have described above, and we emphatically reject the idea that technological interventions can unilaterally "solve" deeply rooted social problems. We explore where and how technical approaches might be *part* of the solution, and how we might exploit their unique properties as a route to broader reforms. In what follows, we propose a series of four potential roles for computing research that may be well-aligned with efforts for broader social change.

The roles we offer here are intentionally modest: they are ways to leverage the particular attributes of computational work without over-claiming its capacity. We argue that these are worthy and plausible aspirations for our nascent field. And like any attempt to effectuate change, the approaches we outline carry hazards of their own, which we also explore below.

Our list is illustrative, not exhaustive, and our categories are not mutually exclusive. In this highly generative moment for technical work oriented toward social problems, we do not seek the last word on how researchers might consider the broader impacts of technical work that responds to social concerns. Rather, recognizing the critiques we have discussed above, our goal is to delineate practical, thoughtful, and engaged paths forward for computing in the service of social change.

11.1 Computing as Diagnostic

Computing helps us understand and measure the contours of social problems.

While computing cannot unilaterally provide solutions to broad social problems, it can yield useful tools to diagnose and precisely characterize those problems. Diagnostic computational approaches, used in tandem with other empirical methods, can provide crucial evidentiary support for work that attends to values in technology — even if computation itself is an insufficient remedy.

Many now-classic studies in the field take this approach, giving us a new sense of the shape and depth of a long-standing problem by applying a computational lens. Latanya Sweeney's analysis of ad delivery platforms demonstrated, among other things, that arrest-related advertisements were more likely to appear in response to searches for first names commonly associated with African-Americans (Sweeney, 2013), likely reflecting Internet users' disproportionate propensity to click on ads suggesting that African-Americans have criminal records. Teams of computer science researchers have demonstrated the gender biases inherent in word embeddings and the difficulties they create for machine translation and other language tasks (Bolukbasi et al., 2016; Caliskan et al., 2017). More recently, Buolamwini and Gebru demonstrated that several commercially available facial analysis systems perform significantly worse on women and individuals with darker skin tones (Buolamwini and Gebru, 2018).

These analyses — and others in the same spirit (e.g. Krauwer, 2003; Obermeyer and Mullainathan, 2019; Raji and Buolamwini, 2019) — have enhanced our ability to document discriminatory patterns in particular socio-technical systems, showing the depth, scope, and pervasiveness of such problems and illuminating the mechanisms through which problems occur. They can provide particular value in contexts in which components of the system are "black-boxed": when the criteria of decision-making (and the values that inhere therein) are obscured by complexity, trade secret, or lack of transparency, diagnostic studies can provide an important means of auditing technical processes (Sandvig et al., 2014). And yet, while these studies sometimes point toward improvements that could be made (e.g., train facial analysis systems on a more diverse data set; use caution with language models trained on text corpora with ingrained human biases), they do not purport to offer computational solutions to the problems they illustrate.

The benefits of computational diagnostics extend beyond socio-technical systems, yielding valuable insights into broader social problems stemming from sources that are not computational in nature. For example, in Chapter 7, we discuss how we can leverage search data to surface unmet health information needs across the African continent, with implications for health policy in the offline world. Researchers have also applied techniques like satellite imagery and spatio-temporal analysis to characterize rural development patterns (Hu et al., 2019; Kulinkina et al., 2016; De-Arteaga et al., 2018). These studies, and many more like them, demonstrate the power that computational techniques have to shed light on a broad range of social issues.

Without purporting to resolve underlying social problems, diagnostic studies that use computational techniques can nevertheless drive real-world impacts on the systems they investigate. For example, in response to Buolamwini and Gebru's findings (Buolamwini and Gebru, 2018), Microsoft and IBM reported that they had improved the accuracy of their facial recognition technologies along gender and racial lines (Puri, 2018; Roach, 2018). Continued advocacy following this work has led to government action questioning, limiting, and prohibiting the use of facial recognition in a number of contexts.

Thus, computational techniques can prove valuable in diagnosing some of the problems raised by socio-technical systems, prompting discussion about the significance of these findings and potential responses. In support of these goals, recent work has sought to advance the diagnostic role that computing can play by providing a principled approach to interrogate the machine learning pipeline (Gebru et al., 2018; Mitchell et al., 2019). These diagnostic efforts do not present themselves as solutions, but rather as tools to rigorously document practices. Thus, when compared with other computing interventions that aim directly at incremental improvements, they are less vulnerable to becoming a substitute for broader change. These efforts are not intended to absolve practitioners of the responsibility to critically examine the system in question, but instead to aid in that investigation.

Some important caveats are warranted. First, computing is not unique in its capacity to help us diagnose social problems. Disciplines like science and technology studies (STS), sociology, and economics provide their own sets of tools to interrogate socio-technical phenomena — including tools that capture important dimensions poorly addressed by computational approaches. For example, descriptive ethnographic research is essential for understanding how social and organizational practices intersect with technical systems to produce certain outcomes (e.g., (Brayne, 2017; Christin, 2018). Computing is only one of multiple approaches that should be brought to bear on the analysis — even when it may appear that the system in question (e.g., facial recognition technology) is primarily technical. A holistic analysis of a socio-technical system must draw from a wide range of disciplines in order to comprehensively identify the issues at stake (Crawford and Calo, 2016; Selbst et al., 2019).

Second, our optimism as to the potential benefits of computing as diagnostic must be tempered by a healthy degree of skepticism. Once metrics and formulae are used to characterize the performance of a system, individuals and organizations may have incentives to optimize towards those metrics in ways that distort their behavior. This process puts any targeted metric at risk of becoming a less useful measure over time — a phenomenon known as Goodhart's Law (Hoskin, 1996) or Campbell's Law (Campbell, 1979). We can see this worry instantiated, for example, in the EEOC's 4/5th rule. The rule applies to employee selection procedures, and holds that when the selection rate for one protected group is less than 80% of that of another group, such a gap is strong evidence of proscribed disparate impact discrimination. Initially intended as a guideline to aid in diagnosis of potentially discriminatory practices, it has become a target in itself for practitioners who strictly enforce the 4/5th rule while sometimes failing to consider other aspects of discrimination and bias. Thus, when using computational techniques as diagnostic tools, care must be taken to prevent diagnostics from implicitly becoming targets.

Third, the diagnostic use of computational techniques requires access to relevant information, and this is a significant barrier in many cases. Analytic efforts can often be stymied or limited by precisely the power structures that could otherwise be diagnosed as in need of change. For example, spurred in part by lobbying by the National Rifle Association, a 1996 spending package included a provision known as the Dickey Amendment that discouraged the Center for Disease Control from studying gun violence in the United States (nra, 1996). A recent federal budget bill clarified that such research is allowable, but declined to appropriate funds for such research (Laslo, 2019; Weixel, 2018). Similar restrictions apply to the study of policing and police shootings — our ability to diagnose the extent of the problem through computational means is hamstrung by the lack of comprehensive data (Davis and Lowery, 2015; Harmon, 2012).

These limitations are also mirrored in the private sector, where corporations operate black-box algorithms, often with little oversight (Pasquale, 2015). In this context, a number of studies seek to characterize proprietary algorithms used in news curation (Lurie and Mustafaraj, 2019), resume search (Chen et al., 2018), and ad delivery (Ali et al., 2019; Datta et al., 2015). While these diagnostic efforts give us insight into the workings of the algorithms they study, it is important to recognize that their methodologies (and very existence) are largely dictated by the means of access afforded by the same corporations they study. In such cases, it can be infeasible to provide a fully comprehensive or conclusive diagnosis.

Finally, it can be tempting to view diagnosis itself as a goal: precisely stating the problem, we might hope, will naturally lead to a solution. Indeed, a good diagnosis can motivate the relevant parties to work towards remedying the concerns it surfaces. The above examples, and others that follow in subsequent sections, demonstrate that illuminating a problem through technical examination can lead toward positive change. However, it is important not to confuse diagnosis with treatment: there are a number of domains in which we are well aware of the extent of the problem (e.g., mass incarceration or homelessness), yet do not have sufficient will or consensus to adequately address it.

11.2 Computing as Formalizer

Computing requires explicit specification of inputs and goals, and can shape how social problems are understood.

People and institutions that make important choices often speak in general terms about what they are doing. A standard that says that social workers must act in the "best interests of the child," for instance, or an employer's stated intent to hire the "most qualified applicant," leave wide latitude for interpretation. This vagueness is a double-edged sword. At its best, such broad flexibility lets decision-makers consider factors that are specific to a particular case or situation, and that could not be stated in a rigid rule. Vagueness is also an important political tool: laws are often worded broadly because no majority of legislators could agree on a more specific plan.[1] Yet an underspecified standard may effectively delegate important choices to people and situations that are less open to scrutiny and supervision, and where individual prejudice may play a larger

[1]In one famous example of legislative imprecision, the Sherman Act – the statute that created and that still governs American antitrust law – forbids "conspiracies in restraint of trade" without making clear what those are.

role. The relative merits of holistic and flexible "standards" versus more rigid and consistent "rules" are a prominent topic of debate among legal scholars.

A computational model of a problem is a statement about how that problem should be understood. To build and use a model is to make, and to promote as useful, a particular lens for understanding what the problem is. The nature of computing is such that it requires explicit choices about inputs, objectives, constraints, and assumptions in a system. Increasing the role of computing within high-stakes decisions, therefore, often moves decision-making away from generalized standards and toward more explicitly specified, rule-based models (Citron, 2018). The job applicant who is ranked first by a statistical model or other algorithm will not simply be the one holistically deemed most qualified; she will be the one ranked first by the operation of discrete numbers and specific rules. Those rules will reflect a range of concrete judgments about how job performance should be defined, measured, and forecasted.

This quality of computing is often bemoaned: algorithms are cold calculators that collapse nuance and execute their tasks in ignorance of the histories and contexts that precede and surround them (Eubanks, 2018a; Noble, 2018; Selbst et al., 2019). But this same quality may also be a source of political potential. Because they must be explicitly specified and precisely formalized, algorithms may help to lay bare the stakes of decision-making and may give people an opportunity to directly confront and contest the values these systems encode. In the best case, the need to disambiguate general policy aims may provide a community with important touch-points for democratic deliberation within policy processes.

Of course, just because decisions must be made about how to formalize a problem does not mean those decisions will necessarily be made in a transparent, inclusive, or accountable way. More often, perhaps, they may not be. Still, the formalism required to construct a mathematical model can serve as an opportune site of contestation, a natural intermediate target for advocacy. Transparency and accountability, moreover, are themselves intermediate virtues: advocates will most often pursue these ends not out of an abstract interest in procedural questions, but rather as a way to influence the substance of decisions that the eventual algorithm will reach.

Complex models are often inscrutable: it may be difficult to describe why particular features were chosen or weighted a certain way. But any such model is surrounded by technical and policy choices that are amenable to broad understanding and debate. The choice of objective function, for instance, and debates over which features to use as candidates for inclusion in the eventual learned model, are frequently delicate and politically important questions. The process of formalization requires that someone make these decisions — and gives us the opportunity to explicitly consider how we would like them to be made.

The growing prominence of computing makes the politics of formalization ever more important. The assumptions in a model may be rigid, but the act of choosing those assumptions is, paradoxically, a creative one, and increasingly important to social and political outcomes. In many settings where automated decision-making tools are used, this creative process of designing a system is an opportunity for non-technical stakeholders, including responsible officials and members of the public, to hash out different ways in which a problem *might* be understood. Questions of this nature include: what is the specific predicted outcome that serves as the basis for decision-making, and why was it chosen over competing alternatives (Passi and Barocas, 2019)? Which factors should be treated as indicia of "good" job performance (Barocas and Selbst, 2016)? How much risk should be considered "high" risk, meriting special treatment or attention from a court or a social services agency (Koepke and Robinson, 2018)?

In some cases, computing researchers have explicitly recognized the formalization process as an opportunity to change how stakeholders perceive the problem domain. For example, researchers who work on matching mechanisms for refugee resettlement have argued that by using more information about local communities — for instance, matching refugees to places that already have a successful cluster of earlier immigrants speaking the same language — they can improve resettlement outcomes and in turn increase the number of refugees that communities are willing to accept (Delacrétaz et al., 2016; Jones and Teytelboym, 2018; Kominers, 2018). Such steps are important because, as Robert Manduca has argued, "many of our pressing social problems cannot be solved by better allocating our existing sets of outcomes" (Manduca, 2019). The process of formalizing problems in this manner can help expand the set of outcomes considered achievable.

In other cases, the formalization process has become a site of political and moral contestation among stakeholders. For instance, using statistical models to assess those who have been accused of crimes is a controversial and widespread practice. These tools are described as "risk assessment" tools, but the probabilities that they actually forecast — such as the chance that the accused person will miss a future appointment related to the case — do not directly correspond to the narrow sense of "risk" that is legally salient at a bail hearing, namely the risk that the defendant will abscond from the jurisdiction or will violently harm another person before a trial can be held (Koepke and Robinson, 2018). In situations like this, where the relevant risk is substantially narrower and more serious than the risks actually being measured and predicted by an algorithm, the algorithm effectively paints with too broad a brush, stigmatizing the accused by implicitly exaggerating the hazard that they pose (Koepke and Robinson, 2018). Many failures to appear at a courtroom hearing are traceable to anodyne causes like a lack of transportation, or confusion about the time and place of the appointment. Formalizing the problem in a way that conflates these risks – as some pretrial instruments do – is a substantive and consequential decision, one that here operates to the detriment of the accused. For this among many other reasons, a broad coalition of more than a hundred civil rights organizations, as well as a substantial and growing number of technical experts on algorithms and criminal law, either oppose the use of these instruments in the forms they currently take or else have raised serious concerns about their use (Leadership Conference on Civil and Human Rights, 2018).

The debate over how to formalize pretrial risk — and over how and indeed whether to use these instruments — is highly active at the time of this writing. Many advocates and scholars have argued that any feasible formalization of the problem, given the available data, will have drawbacks so serious as to counsel against using actuarial methods at all (Eckhouse et al., 2018; Koepke and Robinson, 2018; Leadership Conference on Civil and Human Rights, 2018; Mayson, 2019). Other activists, even while objecting to data-driven methods for these and other reasons, are also seeking to generate new kinds of data for the pretrial system to use, such as formalized and measurable evidence of the community support available to an accused person (Raj Jayadev, 2019). The inherent challenges of formally modelling this problem are driving *both* changes to the models being used, and a reconsideration of whether such models should be used at all.

The moral reckoning of selecting objectives, inputs, and constraints can easily be undertaken without full exploration of alternative possibilities. Without paying attention to the question of whether our formulation carefully reflects the true goals of the domain, the selection of parameters can easily fall prey to inaccurate assumptions or failures of imagination.

For instance, a discussion about the quality of employees in an organization might implicitly assume that the relevant metrics for each employee could be computed as a function of their attributes in isolation. This approach, however, would be unable to assess the quality of the team of employees that results, since a focus on individual attributes would fail to model anything about interactions or complementarities among multiple employees. Such an analysis naturally motivates considerations of synergy and diversity within an organization, via a careful consideration of the underlying objective function (Page, 2008).

Analogously, a discussion about teacher quality might assume that measures of a given teacher's contributions should consider only the student learning that happens inside their own classroom. Such a model would then ignore other ways that teachers are known to contribute to student learning, such as the possibility that the sole teacher of color on an otherwise all-white faculty might serve as an inspiring role model for students of color throughout that school, boosting their results in many subjects (Obama, 2018). School district data is already likely to include information about the gender and race of teachers, and models could be built to consider such data. Building a computational model of teacher performance that considers these factors, where it is possible, would require different choices about the data to be used for model development and the assumed structure of the eventual learned model.

Although moral considerations are often important in determining system design, practical factors — such as the cost savings involved in using readily-available data, rather than gathering or generating different input data — often play a vital role as well (Citron, 2018; Clarke, 1988; Passi and Barocas, 2019). Judgments about the relative importance of moral and practical considerations are themselves both moral and pragmatic in nature, and require a combination of expert knowledge and normative input. Equally importantly, researchers may be able to help illuminate the constraints and biases in existing sources of data — such as child welfare "substantiation" outcomes, official findings of disability, or police arrest patterns — and to underline both the importance and the expense of generating alternative data that may be a necessary precondition for socially responsible forecasting (e.g., (Wald and Woolverton, 1990)). This opportunity for constructive intervention has a corollary — that technical work can draw attention *away* from such issues, if researchers choose to assume that such issues can safely be disregarded, or relegate them to a passing mention.

More broadly, the formalisms that a powerful institution adopts to judge individual people risk focusing moral evaluation and attention on individuals rather than on systems, a trend that some experts have long recognized as invidious. As far back as 1979, Donald Campbell argued that those who study public programs "should refuse to use our skills in ad hominem research. . . . We should be evaluating not students or welfare recipients but alternative policies for dealing with their problems" (Campbell, 1979). More recent work has echoed this call for an evaluative focus on systemic "interventions over [individual] predictions" (Barabas et al., 2017).

Some values may be harder to incorporate into formal models than others, perhaps because they are difficult to quantify (Friedman and Nissenbaum, 1996). What we choose to incorporate into models is also largely driven by the available data, and this constraint may press both scholars and practitioners to rely on measures that elide or distort important aspects of a situation, or that are not conducive to advocating for broader change (Clarke, 1988). For stakeholders primarily committed to values that may be ill-served by the process of formalization, the decision to involve computing in a decision process may itself be a morally controversial one.

11.3 Computing as Rebuttal

Computing can clarify the limits of technical interventions, and of policies premised on them.

Critical reflections on the limits of computing may drive some stakeholders — and at times, the political process as a whole — to reject computational approaches in favor of broader change.

The belief that building and deploying computational tools is a neutral process that can safely be assumed beneficial may be long debunked in academic circles, but it remains a powerful force among policymakers and the general public. When scholars of computing recognize and acknowledge the political valence of their technical work, they can make it more difficult for others to leverage computing — both practically and rhetorically — for political ends.

Technical experts can be particularly effective in contesting claims about technologies' capabilities and neutrality. This opportunity is the converse of the risk that the technical research community may "fail to recognize that the best solution to a problem may not involve technology" (Selbst et al., 2019), and that exploration of technological approaches may distract from broader aims.

For instance, in the context of facial recognition and law enforcement, a growing number of scholars and municipalities have concluded that the best alternative to a bad algorithm might not be a better algorithm — it might be no algorithm at all. Nabil Hassein was (to our knowledge) the first to publicly observe that working to equalize the accuracy of facial recognition across racial groups is in a distinct category of effort from seeking to restrict the use of surveillance based on facial recognition by law enforcement altogether (Hassein, 2017). The former approach addresses specific deficiencies in algorithms used by government actors and others, while the latter argues that the use of even a totally accurate facial recognition algorithm by police would be an amplifier for oppression and injustice.

Earlier, Kevin Haggerty made a similar point about CCTV surveillance. The campaign against it had tried to push back against its adoption by pointing out its lack of effectiveness: "for individuals concerned about the wider social implications of surveillance technologies, debates about evaluation studies are insidious precisely because they fixate exclusively on a standard of 'functioning.' Such a narrow frame ignores the prospect that the true nightmare scenario might be that all of these technologies might actually work; that they might function perfectly, with full enrollment, complete transparency, seamless integration and exacting discriminatory power. Indeed, when critics point out that a surveillance technology does not work, one wonders if they would be thrilled if it did. Rather than confronting the dystopian potentials inherent in the maximum surveillance scenario, the political/methodological knife fight seems to unwittingly help drive systems towards developing ever-greater capacities" (Haggerty, 2009).

Computational researchers focused on the risks of facial recognition seem to have avoided the same trap this round, channeling their insights to raise alarm about both its likely failures and threatening successes. In light of the concerns Buolamwini and Gebru (Buolamwini and Gebru, 2018) raised, U.S. Senator Kamala Harris and others wrote a series of letters to the FBI, FTC, and EEOC requesting information on the legality of facial analysis (Coldewey, 2018). Based on subsequent studies (Raji and Buolamwini, 2019), a number of scholars published an open letter calling on Amazon to stop selling its facial recognition technology to law enforcement altogether (Concerned Researchers, 2019). Several cities in the U.S. have issued bans against police use of facial recognition, with further bans under consideration in the state of California and in the U.K.[2]

Similarly, a group of computing scholars called on Immigration and Customs Enforcement (ICE) to reconsider its plans to use an algorithm to assess whether the applicant would become a "positively contributing member of society" as part of its "Extreme Vetting" program. The experts explained that "no computational methods can provide reliable or objective assessments of the traits that ICE seeks to measure," (Abelson and et. al, 2016) and the program was later abandoned (Harwell and Miroff, 2018). Computing scholars and practitioners can play a critical role in advocating against the inappropriate use of

[2]At the time of this writing, bans have been instituted in San Francisco, Somerville, and Oakland.

computational tools.

Further, by exposing the limits of what a more optimal allocation can accomplish on its own, computing can also help justify the non-computational changes that are necessary for effective reform. A system for allocating scarce housing beds may help give more precise estimates of the housing crisis and shed light on different types of homelessness and what interventions may be more appropriate for each type (Eubanks, 2018a) — which may, in turn, be used to advocate for more progressive policy changes.[3] Or, in the context of child protective services, researchers can make clear that if implementation of a risk assessment algorithm "is accompanied by clarification of goals, analysis of resource needs, increasing training of workers, good supervision, and high-quality services, it will certainly result in a vastly improved system. But without the additional changes, risk assessment will be the emperor's new clothes" (Wald and Woolverton, 1990).

Another aspect of computing as rebuttal is work that uses a formalization of the underlying computational problem to establish mathematically rigorous limits on the power of algorithms to provide certain guarantees. One recent example comes from the study of prediction algorithms for risk assessment: formal analysis has shown that when two groups differ in their base rates for some behavior of interest, any well-calibrated algorithm assigning probabilities of this behavior to individuals must necessarily produce differences between groups in the rate at which people are inaccurately labeled "high-risk" or "low-risk" (Chouldechova, 2017; Kleinberg et al., 2016). This result thus establishes

[3] Eubanks discusses the January 2016 LA homelessness strategy report by Mayor Eric Garcetti which benefited from a previously-implemented coordinated entry system for allocating homelessness services in the city. This report was followed by policy changes that increased funding for low-income housing and homelessness services. (Measure HHH and Measure H.)

that *no matter what* algorithm is employed, any way of assigning risk estimates to two groups of differing base rates will necessarily produce a specific kind of disparity in outcomes between the groups; we cannot eliminate the problem through a better choice of algorithm. Formal results of this type can expose the limitations of an entire category of approaches — in this case, the assignment of numerical risks to individuals — in a way that the deficiencies of any one specific algorithm cannot.

Thus, in a world where practitioners are tempted to pursue computational interventions, technical experts have an inescapably unique contribution to make in illustrating the inherent limits of computational approaches to the task. Those who work on computational approaches can be particularly well positioned to recognize and declare these limits. Computing research has demonstrated the impossibility, infeasibility, or undesirability of proposed policies in other contexts as well. In the area of electronic voting, more than a decade of research and effort has gone into establishing that secure, anonymous, and fully electronic voting systems are infeasible with present and reasonably foreseeable computer science methods. This view is now widely understood by policymakers, and has led large numbers of policymakers away from initial misplaced enthusiasm for paperless high-tech voting equipment. Similarly, an authoritative consensus among computing researchers about the infeasibility of satisfactorily secure "backdoors" in encryption technology has played a major role in dampening policymaker enthusiasm for Orwellian surveillance mandates (Abelson et al., 2015).

Critiques of computing research can also illustrate the limitations of existing policy frameworks. The majority of computational research on fairness is built on frameworks borrowed from discrimination law–for instance, the definition of protected categories, the aforementioned 4/5th rule as a metric for assessing disparate impact, and, perhaps most crucially, the belief that fairness can be achieved by simply altering how we assess people at discrete moments of decision-making (e.g., hiring, lending, etc.). At best, discrimination law is an incomplete mechanism to remedy societies marked by historic injustice and deeply entrenched structures of oppression. Computational fairness work inherits these limitations, and critiques of the narrowness of that work might also be read as critiques of the underlying legal frameworks on which they are built.

11.4 Computing as Synecdoche

Computing can foreground long-standing social problems in a new way.

Virginia Eubanks' acclaimed ethnography *Automating Inequality* tells the stories of three automated systems used to administer public services in different parts of the country (Eubanks, 2018a). We learn about an attempt to automate welfare eligibility determinations in Indiana which wreaks havoc on the lives of the state's most vulnerable. In Los Angeles, the county's "coordinated entry" system ranks homeless individuals in order to apportion the city's limited housing supply — treating basic human needs as resources to be strategically allocated, not rights to be ensured. And in Allegheny County, Pennsylvania, Eubanks tells the story of a risk model used to predict child abuse and neglect that targets poor families for far greater scrutiny.

Eubanks' book draws much-needed attention to these systems. But in discussing the book, and in the text itself, Eubanks is explicit that her core concern is *poverty*, not technology. For Eubanks, computing is just one mechanism through which long-standing poverty policy is manifested. As she terms it, data analytics are "more evolution than revolution": they are a new instance of the tools we've used for generations, built on the same notions of desert and blame that have long characterized poverty discourse. Automated systems sit alongside broader social policies and cultural attitudes that divert poor people from the resources they need. Yet a technological lens does important work for Eubanks: by framing her book through algorithmic instantiations of poverty policy, she brings renewed attention to the plight of the poor writ large. It's safe to say that Eubanks' work has garnered more attention (and attention from different spheres of influence) as a book about inequality through technology than it might have if it were a book about inequality in general.

In this way, computing acts as a synecdoche — a part that stands in for the larger whole in discourse and critique. Computing can offer us a tractable focus through which to notice anew, and bring renewed attention to, old problems. This approach is not uncommon in political discourse. Social problems are by nature complex and multivalent; we rarely obtain purchase on a problem by focusing on its entirety. In discourse and in policymaking, we often chip away at big social issues by concentrating on their components — and for both better and worse, technology critique often captures public attention. Even when we ultimately are concerned about a broader problem and see computing as but one (and perhaps not even the most important) facet of that problem, framing it as a technology problem may be quite pragmatic. Such a focus can leverage resources and attention that might not accrue otherwise. Put most bluntly, many people would not pick up a book about poverty policy in general–but are game to read a critique of the algorithms used to administer it.[4]

Jack Balkin describes this as a *salience* function of new technologies. He proposes that, rather than focusing entirely on technological novelty, researchers should ask: "what elements of the social world does a new technology make particularly salient that went relatively unnoticed before? What features of human activity or of the human condition does a technological change foreground, emphasize, or problematize?" (Balkin, 2004). In his own analysis of the Internet and free speech, Balkin emphasizes that digital technologies "place[] freedom of speech in a new light" and "bring[] features of the system of free expression to the forefront of our concern, reminding us of things about freedom of expression that were always the case" (Balkin, 2004). In writing about Internet speech, Balkin tells us something about characteristics of speech more generally.

The significant risk, of course, is that a focus on the technological aspects of a problem can restrict our attention to merely those aspects. A computing lens can have the effect of masking and pulling political capital away from other and more insidious facets of a problem, as well as other (non-technical) means of addressing it.[5] It can also make computing something of a scapegoat; to reprehend computing for broad social wrongs may give us a convenient target for outrage at systemic injustice, but in a way that does not build momentum toward change. Further, using computing as a synecdoche strategically exploits

[4]As Eubanks writes, the United States stands in "cultural denial" of poverty, "a social process organized and supported by schooling, government, religion, media, and other institutions" on both the left and the right (Eubanks, 2018a) (p. 175).

[5]Eubanks laments this narrowed horizon in the context of poverty policy: "When the digital poorhouse was born, the nation was asking difficult questions: What is our obligation to each other in conditions of inequality? How do we reward caregiving? How do we face economic changes wrought by automation and computerization? The digital poorhouse reframed these big political dilemmas as mundane issues of efficiency and systems engineering: How do we best match need to resource? How do we eliminate fraud and divert the ineligible? How do we do the most with the least money?" (Eubanks, 2018a) (pp. 197-98).

computing's hegemony, but does not, at root, challenge it — and may in fact reinforce tendencies to center the computational dimensions of problems while dismissing their other aspects.

Some of these tensions have recently emerged in critiques of the labor practices that underpin the operation of algorithmic systems. In Finland, prison laborers train a classifier on Finnish-language business articles; the startup running the project paid the prison the equivalent of what it would have paid Mechanical Turk workers for the same tasks. (It is not clear what proportion of the payment went to prisoners themselves.) Though the story was reported predominantly within tech media verticals, and the prison's practices decried by notable critics of technology, others noted that there was "nothing special about AI" (Chen, 2019) in the story: its technological framing was merely a new illustration of long-standing practices. As Lilly Irani noted, "[t]he hook is that we have this kind of hype circulating around AI" as a gloss on "really old forms of labor exploitation." Similar critiques focus on the working conditions of human content moderators who review violent, traumatic imagery for very low wages, and without proper support for the toll such work takes on their mental and physical health (Gillespie, 2018; Roberts, 2019). Low-wage and exploitative work existed long before computing; computing snaps them into focus and exacerbates them, but these are not merely computing problems, and we should not treat them as though they were.[6]

[6]In another labor context, concerns about automation in the workplace (of the "robots will take all the jobs" ilk) may also represent a technological scapegoat for deeper and more pernicious concerns about work precarity, capital accumulation, and the demise of affordable education and the social safety net. The latest waves of workplace automation may be exciting so much concern not because the technology is so fundamentally different from prior advances, but because "the programs that helped Americans deal with how technology is *always* upending the job market were dismantled" (Spross, 2019) (emphasis added). An overemphasis on robots obscures these broader problems and runs the risk of diverting further political capital from them.

Sebastian Benthall makes a pointed critique on this theme. Channeling comments by Luke Stark, Benthall explains that, while debates about technology may at root be debates about capitalism, "it's impolite to say this because it cuts down on the urgency that might drive political action. . . . The buzz of novelty is what gets people's attention" (Benthall, 2018b). Rather than viewing a technology lens as a pragmatic route into a broader social or economic problem, Benthall contends that the AI ethics research community's focus on engineers and tech companies ultimately conflates academic and business practices, creates a convenient "us-versus-them" dialectic (in which the "them" is software developers), and "serves as a moral smokescreen" that impedes effective critique of root problems. Benthall notes, for example, that the emphasis on AI ethics in the critical community is "an awkward synecdoche for the rise of major corporations like Google, Apple, Amazon, Facebook, and Netflix" (Benthall, 2018a) — and that the focus on technology rather than institutionalism or political economy necessarily circumscribes the issues that can be brought to the table.

To recognize that technology is only one component of a broader sociopolitical system is not to give technology a free pass. The existence of a technology may help to create the conditions to support and intensify a particular arrangement of power. Langdon Winner described this quality as technology's inherent compatibility with particular sociopolitical systems (noting, for instance, that nuclear power is inherently compatible with centralized and hierarchical decision-making as the social "operating environment" which allows the technology to function practically, whereas solar energy may be more compatible with distributed egalitarian systems of decision-making) (Winner, 1980). Content moderation systems seem naturally to invite exploitative labor practices based on the scale and speed required and the nature of the content to be moderated; the technology does not just happen to be paired with these practices, but is a close fit for it. A synecdochal focus on computing must walk a pragmatic– and tenuous–line between overemphasis on technical aspects, on the one hand, and due recognition of the work computing does to reinforce social systems, on the other.

11.5 Discussion

Technical work on prediction and optimization of resources in policy domains can sometimes seem to favor slow and incremental approaches to social change — taking a side, without necessarily intending to, against broader and more sweeping redesigns of existing institutional arrangements. But as we have sought to show, technical work can also operate as a constructive ally for broad social change. In this chapter, we have described four ways computing research can support and reinforce more fundamental changes in society and politics. Such research can help to diagnose problems, shape through formalization the ways people understand social problems and possibilities, illuminate the boundaries of what can be achieved by technical means, and make longstanding problems newly salient. And it can do so in ways that are compatible with taking less, rather than more, of today's social landscape for granted.

Of course, while the above framework illustrates different roles that computing research itself can play in social change, it does not explore the full range of activities in which researchers may choose to engage. Computing scholars are well-positioned to clarify the limits of what computing can and cannot do, expose failures in computing systems deployed in social contexts, and identify missed opportunities for computing to assist in broader reforms. Indeed, many of the examples mentioned above go beyond sharing research findings within the academic community. They include efforts to make the work and insights available to those working for social change (e.g., policy makers and activists), and also to build awareness among the broader public. Such engagement from computing scholars can allow us to move towards a human centered design of algorithmic and AI systems and center the voices vulnerable communities (Broussard et al., 2019). These activities are sometimes considered ancillary to the research enterprise, but we believe the question of which activities are or should be counted as "research" is a rich one that deserves further exploration by the community, particularly in the context of efforts for social change.

Part V

Conclusion and Future Work

In this thesis, we discuss ways to build and leverage algorithmic and computational techniques to improve societal welfare. The work presented here draws from and builds on a range of techniques within algorithms and artificial intelligence. The problems we tackle are inspired by a wide-range of domains and are informed by insights from development economics, global health, public policy, science and technology studies (STS), and sociology. As the use of algorithmic and AI techniques becomes more pervasive, there is a growing appreciation of the fact that the most impactful solutions often fall at the interface of various disciplines and an acknowledgement for the potential benefits of a unifying framework for addressing such problems.

We close by outlining a sample of future directions for research at this interface. We present these directions divided up into five general categories – developing nations, education and mentoring, neighborhoods and housing, labor and online platforms, and disadvantage and inequality, more generally. These directions are neither meant to be representative nor exhaustive, but rather to present some potential starting points to explore the use of algorithmic and computational techniques to improve access to opportunity. Many of the problems presented here are borne out of discussions in the MD4SG research community, including in the workshop series, working groups, and colloquium series, as well as a survey (Abebe et al., b; Abebe and Goldner, 2018a,b; Abebe and Goldner).

CHAPTER 12

FUTURE DIRECTIONS

12.1 Access to Opportunity in the Developing World

As discussed in Chapter 7, there are pervasive data inequalities between developing and developed nations. In many settings where we may wish to improve access to information or resources in developing nations, there is a lack of information regarding underlying matters—whether that be the prevalence of diseases or accurate measurements of economic welfare and poverty—due to the unavailability of high-quality, comprehensive, and reliable data (UN, 2014). This limits the implementation of effective policies and interventions.

An emerging solution, which has been successfully demonstrated by the Information Communication Technology for Development (ICTD) research community, has been to take advantage of high phone and Internet penetration rates across developing nations to design new technologies which enable collection and sharing of high-quality data. There has also been recent work from within the AI community to use new data sources to close this information gap, including that presented in Chapter 7. Such AI-driven approaches surface new algorithmic, modeling, and mechanism design questions to improve the lives of many under-served individuals. We discuss two particular areas in further detail below:

12.1.1 Agriculture

Agriculture accounts for a large portion of the economy in many developing nations. Viral disease attacks on crops is a leading cause of food insecurity and poverty. Traditional disease surveillance methods fail to provide adequate information to curtail the impact of diseases (Mwebaze and Biehl, 2016; Mwebaze and Owomugisha, 2016; Quinn et al., 2011). The Cassava Adhoc Surveillance Project implements crowd-sourcing surveillance using pictures taken by mobile phones in order to address this gap (Mutembesa et al., 2018). The tool developed by this Makerere University-based team is set up as a game between farmers and other collaborators, and aims to collect truthful, high-value data (e.g., data from hard-to-reach locations). This approach underlines interesting challenges, such as how to optimally incentivize individuals to collect high-quality information and how to augment this information with existing sources. Note that similar issues have been in other domains—e.g., in citizen science and in computational sustainability (Xue et al., 2016a,b).

Lack of information also leads to inefficiencies in existing systems. For instance, large price discrepancies and major arbitrage opportunities present in markets for agricultural products in Uganda suggest large market inefficiencies (Ssekibuule et al., 2013). To alleviate this, Newman et al. (2018) introduce Kudu—a mobile technology that functions over feature phones via SMS service. Kudu facilitates transactions between farmers in rural areas and buyers at markets in cities by allowing sellers and buyers to post their asks and offers. This contribution, and more broadly concerns around market inefficiencies and information access in developing nations, presents an opportunity to investigate other interventions that can improve outcomes for farmers. For instance, recent work by AgriPredict – an organization working to support small-holder farmers in Zambia – discusses the role of middlemen in markets in Zambia (Simbeye et al., 2019). Modeling the role of middlemen may help identify middlemen-based interventions to improve access to market information for small-holder farmers.

Another prominent inefficiency in agriculture is related to farm lands. Bryan et al. (2017) argue that land fragmentation and misallocation of high productivity land leads to decreased agricultural productivity. Naturally occurring markets are slow at reallocating land to resolve these issues. Bryan et al. (2017) present results from a field experiment in Kenya showing that simple market design-based land exchanges can help alleviate these issues. This work on land reallocation is tied to the fair division literature discussed in Part II and especially in Chapter 5. By bridging this literature gap, and in particular building on the fair division on graphs work, we may be able to identify further opportunities for improving land reallocation that can work under more general studies and may be able to scale to larger problem instances than those studies in Bryan et al. (2017).

12.1.2 Poverty-Reduction

Availability of new technologies also presents opportunities to tackle fundamental problems related to poverty. Advances in last-mile payment technologies, for example, enable large-scale, secure cash transfers. GiveDirectly leverages this and the popularity of mobile money across the world to create a system where donors can directly transfer cash to recipients (GD; Blattman and

Niehaus, 2014). GiveDirectly moves the decision about how to use aid from policy-makers to recipients, giving recipients maximum flexibility. Such aid generates heterogeneity in outcomes—e.g., families may use aid to start a business, pay rent, cover health costs, and so on. Policy-makers used to prioritizing specific outcomes may be uncomfortable by such a model. A research question then is: can we predict how a given population will use aid? Likewise, how can we target people for whom the interventions will make the largest difference? Can we model and analyze the effects of individual- versus community-based targeting? Aid has historically been targeted on the basis of finding the most deprived people. The ability to model heterogeneous treatment effects opens the door for designing more nuanced mechanisms that fairly and efficiently allocate subsidies in order to maximize a desired outcome.

A separate topic of interest is understanding what mechanisms already exist in communities to mitigating the impact of poverty and economic hardship. For instance, across many developing nations, communities use saving circles and their close counterparts. There is a long line of empirical work that studies the role of saving circles within communities and their potential effects (Pankhurst and Mariam, 2000; Ito, 2003; Nnyanja, 2017; Abamagal and Abamagal, 2019). There remain, however, many open questions related to modeling and analyzing the role of saving circles. By modeling these programs, we can ask questions around robustness of existing saving circles, optimal composition of saving circles, and other design questions around their formation. We can also examine the relationship between shocks, such as those studied in Chapter 4 and saving circles.

Problems in the developing world surface unique challenges at the intersection of AI, Information Communication Technology for Development (ICTD), and development economics. Solutions often have to be implemented in resource-constrained environments (e.g., over feature phones or with low network connectivity) (Brunette et al., 2013; Patel et al., 2010). Key populations of interest (e.g., women, people living in rural parts, individuals with disabilities) may not be easily accessible (Sultana et al., 2018; Vashistha et al., 2015b,a). Individuals may have low-literacy (Sambasivan et al., 2010). Lack of understanding of socio-cultural norms and politics, furthermore, may inhibit proposed interventions (Vashistha et al., 2018). All of these highlight the need for a multi-stakeholder approach that leverages technological advances, innovative technical solutions, and partnerships with individuals and organizations that will be impacted by the solutions.

12.2 Education, Information, and Mentoring

There is a long history of research in economics and computation using algorithms and mechanism design-based approaches to improve outcomes in education (Abdulkadiroğlu and Sönmez, 2003; Pathak and Sethuraman, 2011; Abdulkadiroğlu and Sönmez, 1998). Many of these works have helped increase access to quality education for low-income students, students with disabilities, students with special needs, and others belonging to disadvantaged groups.

Within problems related to school choice, outcomes are often evaluated based on the quality of the match of each student. In many settings, however, there may be other desired criteria such as overall diversity of the schools by different demographics. e.g., we may wish to ensure that each school has a fixed percentage of students from a given neighborhood, gender, income-group, or a combination of such factors. How do we best design mechanisms for school choice in settings where are rich diversity requirements? And how can we quantify the change in the quality and composition of matches in these settings?

Information provided to parents, students, teachers, and schools play a crucial role in education. Summarizing a body of work, Chan (2017) argues that information can change outcomes in settings where parents have little information on the costs of colleges or tax credits, on returns to schooling, or on possible available schools. Likewise what information students have regarding college enrollment, selecting courses, or possible careers is a crucial point. Chan (2017) discusses that information-only or nudging-based interventions have been shown to result in a mix of improvements. However, there is less known on the optimal timing, frequency, combination, and model of giving information or providing nudging. Likewise, there remain numerous open questions related to combined effects of providing information and nudges to students as well as their peers, teachers, or parents. Each of these directions leave open questions related to how to best combine and present information by building on insights from education and behavioral economics. Further design questions involve, also, how, whether, and when to democratize this process by allowing students and parents to determine what information or nudges they prefer.

In other questions within education, such as assignment of teachers to schools, desired outcomes may be less well-defined. Assigning teachers to schools is an especially salient problem in many developing nations where there are often centralized processes. Matching questions in this setting are complex. E.g., how do we compare outcomes from assigning high performing teachers to high performing schools or vice versa? Furthermore, how do we tackle these assignment problems when classrooms are heterogeneous, consisting of a mix of high-performing and low-performing students? How do we compare various desirable outcomes in settings where there are different distributions of student performance in classrooms? E.g., do we want to assign higher-performing teachers to class-rooms with a wide-range in student performance or lower-performing teachers where higher performing students may serve as peer instructors to lower-performing students? The availability of both centralized and decentralized markets for assigning teachers to schools also presents opportunities to compare welfare outcomes for teachers, students, and schools (Lo, 2017).

At a high level, the question of setting objective functions or measuring welfare in education systems is a complex one. Based on a survey of related literature, Lo (2017) highlights the challenge of measuring welfare in education systems, asking "what is the underlying objective of education, and how do we measure how well we are doing?"

The study of mentors in education has a long history in the social sciences, ranging from peer-to-peer mentors to teacher-to-student mentors and other forms of mentorship. While there is empirical work that sheds light on what characteristics of mentor-mentee matches appear to improve educational outcomes, there remain numerous questions related to how matches themselves might evolve over time. How do we measure the quality of such matches, when the match quality may itself be multidimensional and dynamic? What aspects of the quality of matches evolve over time and can we predict this? What makes a "good mentor" or a "good mentee"? What aspects of these characteristics are universal and what aspects are dependent on the mentor-mentee matches? The availability of data on mentor-mentee interactions as well as programs through professional societies – such as in computer science and economics – presents an opportunity to tackle some of these questions.

12.3 Communities, Housing, and Homelessness

Allocation of resources—such as public housing, housing vouchers, and homelessness services—has a long history in the economics and computation literature. Even simple-to-state problems in this setting give rise to challenging research questions, many of which are still open. Increased scarcity of housing resources, growing need for services, and the use of algorithmic decision-making tools all open up several avenues with major opportunities for reforming policies and regulations. Below, we discuss some foundational work, new challenges, and opportunities that emerge at the nexus of algorithm and mechanism design, AI, and the social sciences.

Millions of individuals across the US have been evicted or are at risk of experiencing eviction every year. In recent work, Desmond (2012, 2016) shows that eviction is much more common than was previously documented. By compiling the first ever evictions database, Desmond shows that there is an estimate of 2.3 million evictions in 2016 alone and argues that eviction is a leading cause of poverty (EL, 2018). This and related datasets allow us to ask questions related to the causes and consequences of housing instability and eviction on individuals and families. We can, in turn, use these to design algorithms and mechanisms that can improve the allocation of housing services.

In related work, Kube et al. (2018) uses counterfactual predictions to improve homelessness service provisions. By doing so, they realize some gains on reducing the number of families experiencing repeated episodes of homelessness. The Milwaukee Area Renters Study allows us to ask similar questions on what causes eviction. A separate question is related to the impact of eviction on *communities*. Combining data provided through the Eviction Lab and other county-level information, we can also investigate the impact of increasing/decreasing levels of eviction within a community on the fabric of communities and outcomes on labor, education, safety, crime, and other welfare measurements of interest.

At the same time, (Eubanks, 2018a) emphasizes that caution must be taken when using automated decision-making tools for allocating limited resources in such high-stakes scenarios. Eubanks argues that such tools may be used to reduce failure rates by caseworkers; but, if not approached with care, they can deepen already existing inequalities. Furthermore, such tools alone are limited: they do not address the lack of housing and homelessness resources or eliminate human biases or discrimination. It is therefore crucial to take advantage of the confluence of insights from across many disciplines in order to serve the needs of such vulnerable populations.

An issue that is growing in prominence in housing contexts is that of information. Little is documented about how landlords or housing authorities screen applications and make decisions. Ambrose and Diop (2016) show that there is increased restriction in access to rental housing since landlords mitigate information asymmetry by investing in screening tenants. With the increased use and availability of data about individuals, it is of paramount importance to understand the role of information in the decision-making process of entities, such as landlords or housing agencies, who have enormous discretion in how and whether families are housed.

Due to increased scarcity of resources, allocation protocols often involves waiting lists and priority groups. Policy constraints make wait-list design a dynamic rationing problem rather than the static assignment problem discussed above. Dynamic mechanisms present several technical and practical challenges; e.g., incentive-compatibility may be infeasible in dynamic settings due to waiting time trade-offs for applicants. There are consequential design decisions related to how to manage waitlists and different metropolitan areas have different policies (e.g., setting priority groups, conditions under which individuals are removed from the waiting list, set of choices, and many others). Each of these policies impacts the dynamics of the allocation process, waiting time, and quality of matches. Recent work has studied how to design mechanisms satisfying various desiderata and quantify differences in quality of matches across various mechanisms (Arnosti and Shi, 2018; Leshno, 2017; Thakral, 2016; Waldinger, 2017).

A crucial question related to the above is the use of algorithmic tools to screen potential tenants. Despite the pervasive use of such tools, especially in large cities in the US, there is little known about what data these tools use, how and whether possible algorithmic discrimination is accounted for in the screening process, and how land-lords use information recommendations provided by vendors that provide these tools. As the use of such tools is expanding, it is of paramount importance to ask such questions in parallel.

12.4 Labor, Online Platforms, and Discrimination

Online platforms are ubiquitous, providing a vast playground for algorithm design and artificial intelligence. Every policy decision, however, impacts and interacts with the platform's strategic users and there is increased attention being paid to labor as mediated by online platforms and potential discrimination. Past work has begun to investigate some aspects of platforms, of strategic agents, and of discrimination in labor markets, but there are still major opportunities for work at the intersection.

One central issue surrounding labor markets is that of *hiring*, in which a firm takes information about a potential candidate and makes an employment decision. Firms act as classifiers, labeling each applicant as "hire" or "not hire" based on an applicant's "features," such as educational investment or a worker's productivity reputation. In the process of making hiring decisions, however, the firm may potentially make discriminatory decisions—perhaps by using protected attributes, or by not correcting for differences in applications that stem from systemic discrimination (Bertrand and Mullainathan, 2004; Marlowe et al., 1996). Bias in hiring decisions may arise due to implicit human bias or algorithmic bias, in which algorithms replicate human and/or historic discrimination that is reflected in the data on which they are trained (Broussard, 2018; Eubanks, 2018a; Noble, 2018; O'neil, 2016).

One recent line of work investigates hiring policies that achieve diversity or statistical parity (with respect to certain groups) among the hired workers, and how workers make their investment decisions (e.g. whether to attend college) based on the hiring policies in place. Coate and Loury (1993); Fryer Jr and Loury

(2013); Hu and Chen (2017) study settings where there is some known underlying bias or historical discrimination against certain groups; the aim is to characterize hiring policies that are optimal-subject-to-fair-hiring, and to quantify any loss in efficiency compared to optimal-but-discriminatory policies. These works explore two settings: first, when hiring decisions must be "group-blind," that is, they cannot take group membership into account, and second, when they are "group-aware". The aim is to choose hiring policies that will mitigate discrimination against protected categories. Hu and Chen (2017) highlights additional complexity that arises in dynamic settings where workers are hired based on investment decisions (e.g. college GPA) in an initial temporary labor market (e.g. internships) and this job creates a worker's initial productivity reputation that is then used in the permanent labor market. Many of these findings also discuss "trade-offs" between group-blind and group-aware policies.

Another aspect of labor markets is that a worker may have the ability to pay to change a feature of her application in some illegitimate or unfair way in order to improve her outcome in the labor market. Hardt et al. (2016) examine this problem from a robust machine learning perspective. Under certain assumptions of the cost required to change an applicant's reputation, they characterize classifiers that optimally compared to the original reputation (before the applicant modified it).

These are only two aspects at the interplay between hiring and strategic agents; hiring, furthermore, is only one aspect of the labor market. Consider today's popular online labor markets, such as Mechanical Turk, Upwork, Task Rabbit, and Lyft, in which the platform's goal is to match workers to employers or jobs. In these labor markets, the platform's decisions, even at a granular level, have a large impact on the workers and firms. Consider the following platform decisions. Visibility: how many firms can workers see at a time? What capacity do they have to search for job offers? Can workers see jobs and jobs see workers? Initiation: which side (or both) can submit applications? Initiate messaging? Set contract terms? Information: what information is displayed about parties on the opposite side? Name? Photo? Ethnicity? Wage history? Reputation?

Each of these decisions impacts the outcome—not only the quality of the match, but also whether (and how much) discrimination occurs. In a recent paper, Levy and Barocas (2017) outline categories of platform decisions which may mitigate or perpetuate discrimination in labor markets, including the high-level categories of setting platform discrimination policies or norms, structuring information and interactions, and monitoring/evaluating discriminatory conduct.

In offline labor markets, it may be challenging or infeasible to collect data to understand the nature and extent of discrimination. Online labor markets, on the other hand, yield rich data about employer-employee interactions and present the possibility of conducting experiments aimed at reducing bias and discrimination or other desired societal objectives. For instance, Barach and Horton (2017) look at the impact on hiring of hiding workers' wage history. Horton and Johari (2015); Horton (2018) look at the impact of trying to elicit additional information (features) from workers or firms, and the impact of this strategically-reported information on hiring. Horton (2017) examines who the hired worker population is when a minimum wage is imposed on one platform. Each of these provide insights into labor dynamics that may inform platform design and interventions.

12.5 Disadvantage and Inequality

Much of the work presented in this thesis, whether it is related to health, housing, or poverty, is inspired by the fundamental observation that many forms of disadvantage are multi-faceted, dynamic, and difficult to measure (Grusky, 2018; Grusky and Weeden, 2016; Tumin et al., 2012). In the very first chapter on "the questions we ask about inequality," in his book *Social Stratification*, Grusky taxonomize asset groups – economic, power, cultural, social, honorific, civil, human, and physical – based on work by Grusky and Weisshaar (Grusky, 2018). This article further highlight that each of these can present themselves in many different forms. For instance, economic disadvantage can manifest itself through wealth, income, or ownership. Likewise, cultural advantage can manifest itself through knowledge, digital culture, and "good" manners.

Despite this long line of work in the social sciences about the many forms of disadvantage, interventions often focus on a narrow measurement of welfare. We discuss this in the context of economic welfare in Chapter 4. With the availability of large datasets and increasing collaboration across disciplines, there is the prospect that we may be able to bring new insights into some of these under-explored facets of disadvantage, provide measurements that adequately capture complex interactions yet are simple enough to inform interventions, as well as bring light to new forms of disadvantage. An example presented in this thesis is measurements around access to quality health information.

One promising direction is related to networks and inequality. Algorithmic and computational work has developed a large set of techniques related to social and information networks that has allowed us to ask questions around how information spreads, opinions form, polarization emerges, and networks evolve. Questions around disadvantage and inequality can bring forth new network-based questions. Some examples include: how can we quantify social capital, how it forms, and how it evolves? What relationships exist between the distribution of individual welfare (such as income distributions) and global measurements of network health (such as integration)? How do networks evolve as a response to, for instance, large positive or negative shocks to a member? What characterizes networks predominantly comprising disadvantaged communities? How do state-of-the-art algorithms in networks – for instance for link prediction – perform for advantaged versus disadvantaged communities? What are the societal and policy implications for these?

APPENDIX A

SUBSIDY ALLOCATIONS IN THE PRESENCE OF INCOME SHOCKS

We present missing proofs and examples from the main text in this section.

A.1 Non-zero Wealth: The Exponential Case

We consider three forms of subsidies – a wealth subsidy, income subsidy, and mixed income and wealth subsidy – in this setting. Below, we show how to optimally allocate subsidies in each of the three settings.

Wealth Subsidies. A wealth subsidy of z_i reduces the ruin probability from $\psi(c_i, u_i, \beta_i, \delta_i)$ to $\psi(c_i, u_i + z_i, \beta_i, \delta_i)$. For the min-max objective, we can proceed exactly as in the previous section, applying the priority algorithm from Section 4.2.1 using the functions $f_i(z_i) = \psi(c_i, u_i + z_i, \beta_i, \delta_i)$, and hence ordering agents by their ruin probabilities.

For the min-sum objective, we use partial derivatives as in the previous section as well. Specifically, we define our objective function to be

$$\gamma(z_1, \ldots, z_n) \quad = \quad \sum_{i=1}^{n} w_i \psi(c_i, u_i + z_i, \beta_i, \mu_i) = \sum_i w_i \frac{\beta_i}{c_i \delta_i} e^{\left(\frac{\beta_i}{c_i} - \delta_i\right)(u_i + z_i)}.$$

We denote the term associated with z_i with γ_i. Taking the partial derivative with respect to z_i, we get

$$\frac{\partial \gamma}{\partial z_i} = \left(\frac{\beta_i}{c_i} - \delta_i\right) \frac{w_i \beta_i}{c_i \delta_i} e^{\left(\frac{\beta_i}{c_i} - \delta_i\right)(u_i + z_i)}. \tag{A.1}$$

Since the process has positive drift, all of the partial derivatives are negative. It is easy to verify that γ is strictly convex since the second derivative with respect to z_i is strictly positive. We therefore define $f_i(z_i) = -\frac{\partial \gamma}{\partial z_i}$ and use this as a priority ordering for the algorithm from Section 4.2.1.

Income Subsidies. The case for the income subsidy proceeds in the same way as the wealth subsidy solution above. We optimize for the min-max objective using the ruin probabilities. For the min-sum objective function, we define

$$\phi(x_1, \ldots, x_n) \quad = \quad \sum_{i=1}^{n} w_i \psi(c_i + x_i, u_i, \beta_i, \mu_i) = \sum_i w_i \frac{\beta_i}{(c_i + x_i)\delta_i} e^{\left(\frac{\beta_i}{(c_i+x_i)\delta_i} - \delta_i\right) u_i}.$$

We take the partial derivative with respect to x_i to get

$$\frac{\partial \phi}{\partial x_i} = -\frac{w_i \beta_i \, \mathrm{e}^{-u_i\left(\delta_i - \frac{\beta_i}{c_i + x_i}\right)}}{\delta_i \, (c_i + x_i)^2} - \frac{\beta_i^2 \, u_i \, \mathrm{e}^{-u_i\left(\delta_i - \frac{\beta_i}{c_i + x_i}\right)}}{\delta_i \, (c_i + x_i)^3}. \tag{A.2}$$

We again has a convex minimization problem and we use $f_i(x_i) = -\frac{\partial \phi}{\partial x_i}$ as an ordering for the priority algorithm to find the optimal solution.

Mixed Income and Wealth Subsidies. We finally consider the case where the planner can give a mix of income and wealth subsidies. We assume that each unit of income subsidy counts toward the budget at a factor of k times the contribution of each unit of wealth subsidy. So, we have a total budget of $B_u + kB_c = B$, where B_c is the total income subsidy and B_u is the total wealth subsidy allocated.

The optimal solution here directly applies the solutions for optimally allocating income and wealth subsidies shown above. Namely, our solution involves $2n$ variables corresponding to an income and wealth subsidy variable for each of the n agents.

As above, for the min-max objective, we can proceed by applying the priority algorithm using the functions, $f_i(z_i) = \psi(c_i, u_i + z_i, \beta_i, \delta_i)$ and $g_i(x_i) = \psi(c_i + x_i, u_i, \beta_i, \delta_i)$ and hence ordering the agents by their ruin probabilities. We abide by the constraint that $\sum_i kx_i + \sum_i z_i$ is at most B.

As for the min-sum objective, we define $f_i(z_i) = -\frac{\partial \gamma}{\partial z_i}$ from Equation A.1 and $g_i(x_i) = -\frac{\partial \phi}{\partial x_i}$ from Equation A.2. We use these f_i, g_i in our priority algorithm, thus ordering the agents by both of the partial derivatives.

A.1.1 Contrasting Prioritizations

Using the case where shocks are drawn from an exponential distribution, we can note a rich set of examples where prioritizations by income, wealth, and ruin probabilities can vary substantially from our solution for the income and wealth subsidy problems.

Lemma 6. The priority ordering given by ruin probability (and likewise the wealth subsidy solution) can be the reverse of the priority ordering given by income and wealth, even when the priority ordering by income and wealth agree.

Proof. Given an agent i, we set $u_i, c_i = 1 + i\epsilon$ and $\delta_i = 1/c_i$. The priority orderings by income and wealth agree and have ordering $(1, 2, \ldots, n)$.

On the other hand, we note that the ruin probability for each agent i is given by:

$$\frac{\beta_i}{c_i \delta_i} e^{\left(\frac{\beta_i}{c_i} - \delta_i\right) u_i} = \frac{\beta_i}{e^{(1-\beta_i)}}.$$

We set $\beta_i = \frac{1}{n-i+2} - \epsilon$. Note that $\beta_i \in (0, 0.5)$. As i increases, the ruin probability above also increases. Therefore, the priority ordering given by the ruin probability is the exact reverse of the orderings given by the agents' income and wealth.

To note a similar result for the wealth subsidy, we first recall that each agent's value for Equation A.1 is given by

$$\left(\frac{\beta_i}{c_i} - \delta_i\right)\frac{\beta_i}{c_i\delta_i}e^{\left(\frac{\beta_i}{c_i}-\delta_i\right)u_i} = \left(\frac{\beta_i^2 - \beta_i}{c_i}\right)e^{\beta_i - 1}$$

This equation has root $\beta_i = \frac{\sqrt{5}-1}{2}$. Thus, setting the β_i as above, we note that this value decreases as i increases, giving us a priority ordering that coincides with the ordering given by the ruin probability. □

Lemma 7. The priority ordering given by the ruin probability can be the reverse of the priority ordering given by the optimal solution for income subsidy and wealth subsidy, even when the latter two coincide.

Proof. We let $\delta_i, u_i = 1$. For notational convenience, we set $r_i = \beta_i/c_i$. The ruin probability is given by

$$r_i e^{r-1}.$$

Equation A.1 yields

$$(r_i - 1)r_i e^{r_i - 1}.$$

Note again that this has root $r_i = \frac{\sqrt{5}-1}{2}$.

On the other hand, Equation A.2 yields

$$-\frac{\beta_i e^{\frac{\beta_i^2}{c_i}-1}}{c_i^2} - \frac{\beta_i e^{\frac{\beta_i^2}{c_i}-1}}{c_i^3}.$$

We set $\beta_i = \left(\frac{\sqrt{5}-1}{2} + \epsilon i\right)c_i$ and let $c_i = 1 - \frac{1}{i+1} - \epsilon$. Therefore, $r_i = \frac{\sqrt{5}-1}{2} + \epsilon i$. We note that the ruin probabilities are increasing in i.

On the other hand Equations A.1 and Equation A.2, used to determine the priority ordering for wealth and income subsidies are both increasing in i, giving us a priority ordering that is the reverse of that obtained by using the ruin probabilities. □

Lemma 8. The priority ordering given by the optimal solution for income subsidy can be the reverse of the priority ordering given by the optimal solution for the wealth subsidy.

Proof. For each agent i, we let $c_i, u_i, \delta_i = 1$. Evaluating Equation A.1, we obtain

$$\frac{(\beta_i - 1)\beta_i e^{\beta_i}}{e}.$$

And, evaluating Equation A.2, we obtain

$$-\beta_i e^{\beta_i - 1} - \beta_i^2 e^{\beta_i - 1}.$$

Note that in the region $\beta_i \in [0.7, 0.8]$, the first value is increasing while the second value is decreasing. Therefore, if we set $\beta_i = 0.7 + \frac{0.1i}{n}$, for each agent i, the optimal solution for wealth subsidy would prioritize agents with smaller indices i and the optimal solution for income subsidy would prioritize agents with higher indices. □

A.1.2 Monotonicity Results

We can establish the monotonicity results as follows. For Equation A.1 evaluated at $z_i = 0$, we get

$$\left(\frac{\beta_i}{c_i} - \delta_i\right)\psi(c_i, u_i, \beta_i, \delta_i).$$

Since u_i only shows up in ruin probability, which is monotone in wealth, this above expression is also monotone in wealth. Likewise, Equation A.2 evaluated at $x_i = 0$ gives us:

$$-\left(\frac{w_i \beta_i \, \mathrm{e}^{-u_i\left(\delta_i - \frac{\beta_i}{c_i}\right)}}{\delta_i {c_i}^2} + \frac{\beta_i^2 \, u_i \, \mathrm{e}^{-u_i\left(\delta_i - \frac{\beta_i}{c_i}\right)}}{\delta_i {c_i}^3}\right).$$

It is easy to see monotonicity by β_i, c_i, and δ_i.

We show non-monotonicity results using the following example with three agents:

Example 52 (Wealth Subsidy)**.** We first show that Equation A.1 is not monotone in c_i, β_i, δ_i. Suppose that $w_i, c_i, u_i, \delta_i = 1$ and $\beta_i = 0.5$ for each agent. Then, the value for Equation A.1 is -0.1516. To see non-monotonicity, we change the values for c_i, β_i, δ_i as follows and evaluate Equation A.1:

c: Set $c = (0.6, 1, 2)$, which evaluate to $(-0.1176, -0.1516, -0.0886)$.

β: Set $\beta = (0.1, 0.5, 0.9)$, which evaluate to $(-0.037, -0.152, -0.081)$.

δ: Set $\delta = (0.6, 1, 2)$, which evaluate to $(-0.0754, -0.1516, -0.0837)$.

Each of these induce a priority ordering $(2, 1, 3)$ for the wealth subsidy instead of the ordering $(1, 2, 3)$ that would be implied by monotonicity.

We can show non-monotonicity in u_i for Equation A.2 using the following example with three agents.

Example 53 (Income Subsidy)**.** Suppose we have agents where $w_i, c_i, \delta_i = 1$ and $\beta = 0.7$. Set their wealth to be $(0.1, 2, 5)$. Equation A.2 evaluates to $(-0.7269, -0.9220, -0.7029)$, giving us the claimed non-monotonicity result.

APPENDIX B

A TRUTHFUL MECHANISM FOR ONE-SIDED MATCHING

B.1 The Lower Bound Construction

The construction uses a collection of overlapping submarkets named $M_0, M_1, \ldots, M_s$. Associated with these markets are integer parameters k_r, for $1 \leq r \leq s$. There will be k_r copies of M_r in the construction. As we shall see, $k_s = 1$, thus there will be exactly one copy of M_s.

Next, in Figure B.1 and Table B.1, we show the form of M_0. The nodes in this figure correspond to the items, and a directed edge (α, β), labeled by the name of an agent, indicates that this agent was allocated portions of item α in the initial solution and portions of β in the final solution. Items and bidders occur with multiplicity possibly greater than 1 and this is called their size. In the initial equilibrium, every item is fully allocated as the total size of the bidders and items are the same; in the final equilibrium, item A_0 is the one item that is not fully allocated. Note that in each equilibrium, the conditions (5.3)–(5.5) from Section 5.5.2 are satisfied.

We continue by presenting the constructions of markets M_r, for $1 \leq r < s$ and of market M_s in Tables B.2 and B.3, respectively. M_s is very similar to M_r; the only difference lies in the presence of one additional item I_s, which is the item e_s, the losing bidder, will receive in the final equilibrium, plus one additional bidder, i_s, who leaves in the final setting. In these markets, all items are fully allocated in both equilibria.

To complete the construction we have to show that the various unspecified

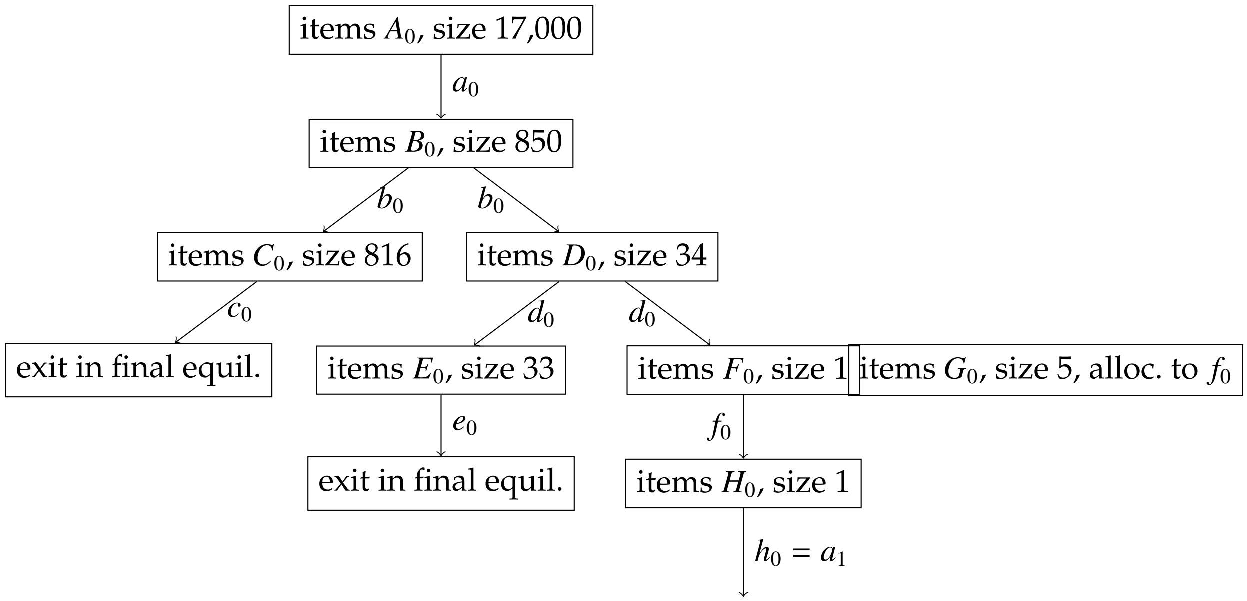


Figure B.1: The allocation in Market M_0.

parameters can be chosen so that the conditions of (5.5) are satisfied for every item-bidder pair. (It is immediate that (5.3)–(5.4) are satisfied.)

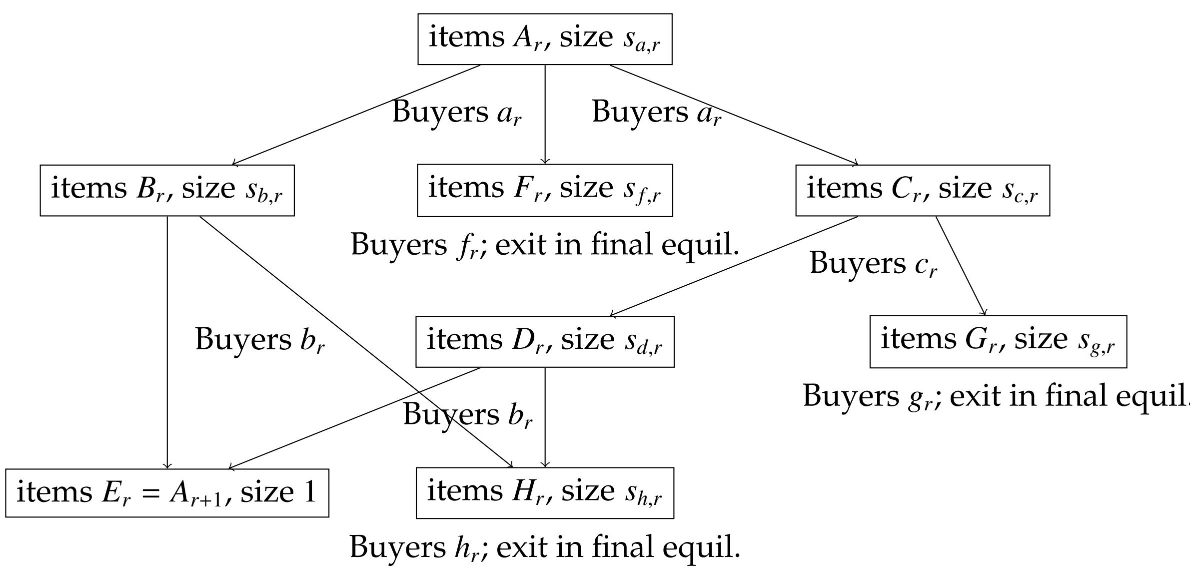


Figure B.2: The allocation Market M_r, $1 \leq r < s$. The parameters are specific in the proof of Lemma 54.

Lemma 54. *There are choices of values for the unspecified parameters for which the*

item		A_0	B_0	C_0	D_0	E_0	F_0	G_0	$H_0 = A_1$	q value
size		17,000	850	816	34	33	1	5	1	
The initial equilibrium:										
Bidder	size									0
a_0	17,000	**1**	1.5	0	0	0	0	0	0	1
b_0	850	0	**1**	$\frac{4687}{7008}$	1.5	0	0	0	0	0.5
c_0	816	0	0	**1**	0	0	0	0	0	0
d_0	34	0	0	0	**1**	5/11	2	0	0	0
e_0	33	0	0	0	0	**1**	0	0	0	0
f_0	6	0	0	0	0	0	$\mathbf{\frac{8}{3}}$	$\mathbf{\frac{2}{3}}$	$\frac{5}{3}$	$\frac{2}{3}$
h_0	1	0	0	0	0	0	0	0	**1**	0
t value		0	0.5	1	1	1	2	0	1	
The final equilibrium:										
a_0	17,000	$\mathbf{\frac{40}{41}}$	$\mathbf{\frac{60}{41}}$	0	0	0	0	0	0	$\frac{40}{41}$
b_0	850	0	$\frac{292}{205}$	$\mathbf{\frac{4687}{4920}}$	$\mathbf{\frac{438}{205}}$	0	0	0	0	$\frac{192}{205}$
d_0	34	0	0	0	2	$\mathbf{\frac{10}{11}}$	**4**	0	0	$\frac{4}{5}$
f_0	6	0	0	0	0	0	$\frac{16}{5}$	$\mathbf{\frac{4}{5}}$	**2**	0
$h_0 = a_1$	1	0	0	0	0	0	0	0	2	0
t value		0	$\frac{20}{41}$	$\frac{79}{4920}$	$\frac{6}{5}$	$\frac{6}{55}$	$\frac{16}{5}$	$\frac{4}{5}$	$2 = t_1^{A,F}$	

Table B.1: Market M_0, showing normalized valuations, multiplicity of bidders and items (their sizes), assignments (in bold), and the t and q values, for both the initial and final equilibria. The overlap with Market M_1 lis in item H_0 which is also item A_1 and bidder h_0 who is also bidder a_1. Note that the t values for $H_0 = A_1$ are the same in markets M_0 and M_1 in both equilibria, as are q values for $h_o = a_1$.

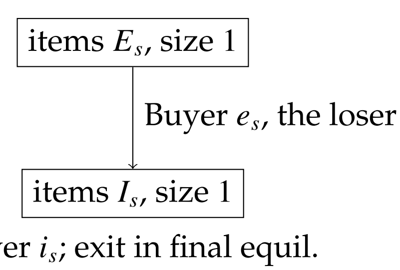


Figure B.3: The allocation in Market M_s.

item		A_r	B_r	C_r	D_r	E_r $= A_{r+1}$	F_r	G_r	H_r	q value
size		$s_{a,r}$	$s_{b,r}$	$s_{c,r}$	$s_{d,r}$	1	$s_{f,r}$	$s_{g,r}$	$s_{h,r}$	
Initial equilibrium:										
Bidder	size									
a_r	$s_{a,r}$	**1**	2	$\frac{1}{2}$	0	0	$v^I_{f,r}$	0	0	0
b_r	$s_{b,r} + s_{d,r}$	0	$\mathbf{\frac{16}{7}}$	0	$\mathbf{\frac{2}{7}}$	$\frac{9}{7}$	0	0	$v^I_{h,r}$	$\frac{2}{7}$
c_r	$s_{c,r}$	0	0	**1**	$\frac{1}{2}$	0	0	$v^I_{g,r}$	0	$\frac{1}{2}$
f_r	$s_{f,r}$	0	0	0	0	0	**1**	0	0	0
g_r	$s_{g,r}$	0	0	0	0	0	0	**1**	0	0
h_r	$s_{h,r}$	0	0	0	0	0	0	0	**1**	0
t value		1	2	$\frac{1}{2}$	0	1	1	1	1	
Final equilibrium:										
a_r	$s_{a,r}$	$\boldsymbol{\frac{8}{9}v_r}$	$\boldsymbol{\frac{16}{9}v_r}$	$\boldsymbol{\frac{4}{9}v_r}$	0	0	$\boldsymbol{v^F_{f,r}}$	0	0	0
b_r	$s_{b,r} + s_{d,r}$	0	$\frac{16}{9}v_r$	0	$\frac{2}{9}v_r$	$\boldsymbol{v_r}$	0	0	$\boldsymbol{v^F_{h,r}}$	0
c_r	$s_{c,r}$	0	0	$\frac{4}{9}v_r$	$\boldsymbol{\frac{2}{9}v_r}$	0	0	$\boldsymbol{v^F_{g,r}}$	0	0
t value		$\frac{8}{9}v_r$	$\frac{16}{9}v_r$	$\frac{4}{9}v_r$	$\frac{2}{9}v_r$	v_r	$v^F_{f,r}$	$v^F_{g,r}$	$v^F_{h,r}$	

Table B.2: Component C_r, showing normalized valuations, multiplicity of bidders and items (their sizes), and the t and q values, for both the initial and final equilibria. $v_r^{a,F}$ and $v_r^{c,F}$ are normalized factors, equal to the value of the assignments using the initial valuations.

item		E_s	I_s	q value
size		1	1	
The initial equilibrium:				
Bidder	size			
e_s	1	**1**	$1/(v_s + 1)$	0
i_s	1	0	**1**	0
t value		1	1	
The final equilibrium:				
e_s	1	$v_s + 1$	1	1
t value		v_s	0	

Table B.3: Component C_s, showing the additional portion in addition to the part shown in the preceding table.

valuations specified above yield the claimed initial and final equilibria.

Proof. We need to choose the values $v^I_{f,r}$, $v^I_{g,r}$, $v^I_{h,r}$, $v^F_{f,r}$, $v^F_{g,r}$, $v^F_{h,r}$, the sizes k_r, and the size proportionality factors $s_{a,r}$, $s_{b,r}$, $s_{c,r}$, $s_{d,r}$, $s_{f,r}$, $s_{g,r}$, $s_{h,r}$ so that for each buyer in each equilibrium, its average value is 1, for $1 \le r \le s$. First, we set $s_{a,r} = s_{b,r} + s_{f,r} + s_{c,r}$, $s_{c,r} = s_{d,r} + s_{g,r}$, and $s_{b,r} + s_{d,r} = 1 + s_{h,r}$.

Next, we observe that because the buyer $h_0 = a_1$, its values for $H_0 = A_1$ are the same, i.e., $2 = \frac{8}{9}v_1$, or $v_1 = 2 \cdot \frac{9}{8}$. Similarly, item $E_r = A_{r+1}$, so $v_{r+1} = \frac{9}{8}v_r$. We conclude that $v_r = 2 \cdot \left(\frac{9}{8}\right)^r$.

Now, for buyer b_r, we choose $s_{b,r} = \frac{5}{9}s_{d,r}$, for

$$\frac{\frac{16}{7}s_{b,r}k_r + \frac{2}{7}s_{d,r}k_r}{s_{b,r}k_r + s_{d,r}k_r} = \frac{\frac{16}{7} \cdot \frac{5}{9} + \frac{2}{7}}{\frac{5}{9} + 1} = 1.$$

Thus $s_{d,r} = \frac{9}{14}(1 + s_{h,r})$ and $s_{b,r} = \frac{5}{14}(1 + s_{h,r})$. To ensure these values are integers, we will make sure that $1 + s_{h,r}$ is an integer multiple of 14.

We turn to the values $v^F_{f,r}$, $v^F_{g,r}$, $v^F_{h,r}$. We choose $s_{h,r} = 14\lfloor v_r/14 \rfloor + 13$, and $v^F_{h,r}$ to satisfy $v^F_{h,r}s_{h,r} + v_r = s_{h,r} + 1$; i.e., $v^F_{h,r} = (\lfloor v_r \rfloor + 1 - v_r)/\lfloor v_r \rfloor$. We need to confirm that $v^I_{h,r} \le \frac{9}{7}$; but $v^I_{h,r} = 9/(7v_r) \cdot v^F_{h,r} < v^F_{h,r} < 1$, as $v_r \ge \frac{9}{4}$.

Similarly, when $v_r > \frac{9}{2}$, we set $s_{g,r} = \lfloor \frac{2}{9}v_r \rfloor s_{d,r}$ (for when $v_r \le \frac{9}{2}$, this would set $s_{g,r} = 0$), and $v^F_{g,r} = s_{g,r} + \frac{2}{9}v_r s_{d,r} = s_{g,r} + s_{d,r}$; i.e., $v^F_{g,r} = (\lfloor \frac{2}{9}v_r \rfloor s_{d,r} + s_{d,r} - \frac{2}{9}v_r s_{d,r})/\lfloor \frac{2}{9}v_r \rfloor s_{d,r}$. Again, we need to confirm that $v^I_{g,r} \le \frac{3}{2}$; but $v^I_{g,r} = 9/(4v_r) \cdot v^F_{g,r} \le v^F_{g,r} \le 1$, as $v_r > \frac{9}{2}$.

When $v_r < \frac{9}{2}$, we set $s_{g,r} = \lceil \frac{9-2v_r}{\frac{2}{3}v_r - 1} \rceil$ and $v^F_{g,r} = (s_{g,r} + s_{d,r} - \frac{2}{9}v_r s_{d,r})/s_{g,r}$; but then $s_{d,r} = 9$, so $v^F_{g,r} = 1 + (9 - 2v_r)/\lceil \frac{9-2v_r}{\frac{2}{3}v_r - 1} \rceil$. Again, we need to confirm that $v^I_{g,r} \le \frac{3}{2}$; but $v^I_{g,r} = 9/(4v_r) \cdot v^F_{g,r} \le 9/(4v_r)(1 + \frac{2}{3}v_r - 1) = \frac{3}{2}$.

As $v_r = 2 \cdot \left(\frac{9}{8}\right)^r$, $v_r \ne \frac{9}{2}$ for any r.

Also, we set $s_{f,r} = \lfloor \frac{16}{9} v_r \rfloor s_{b,r} + \lfloor \frac{4}{9} v_r \rfloor s_{c,r}$ and $v^F_{f,r} s_{f,r} + \frac{16}{9} v_r s_{b,r} + \frac{4}{9} v_r s_{c,r} = s_{f,r} + s_{b,r} + s_{c,r}$; i.e., $v^F_{f,r} = (\lfloor \frac{16}{9} v_r \rfloor s_{b,r} + \lfloor \frac{4}{9} v_r \rfloor s_{c,r} + s_{b,r} + s_{c,r} - (\frac{16}{9} v_r s_{b,r} + \frac{4}{9} v_r s_{c,r}))/(\lfloor \frac{16}{9} v_r \rfloor s_{b,r} + \lfloor \frac{4}{9} v_r \rfloor s_{c,r})$.

We can now calculate the following values.

$$
\begin{aligned}
s_{h,r} &= 14\lfloor v_r/14 \rfloor + 13 \\
s_{d,r} &= 9(\lfloor v_r/14 \rfloor + 1) \\
s_{b,r} &= 5(\lfloor v_r/14 \rfloor + 1) \\
s_{g,r} &= 9\lfloor 2v_r/9 \rfloor (\lfloor v_r/14 \rfloor + 1) \ \text{ for } v_r > \frac{9}{2} \\
s_{g,r} &= \lceil \frac{9 - 2v_r}{\frac{2}{3} v_r - 1} \rceil \ \text{ for } v_r < \frac{9}{2} \\
s_{c,r} &= 9(\lfloor v_r/14 \rfloor + 1) \cdot (\lfloor 2v_r/9 \rfloor + 1) \ \text{ for } v_r > \frac{9}{2} \\
s_{c,r} &= 9 + \lceil \frac{9 - 2v_r}{\frac{2}{3} v_r - 1} \rceil \ \text{ for } v_r < \frac{9}{2} \\
s_{f,r} &= 5\lfloor 16v_r/9 \rfloor (\lfloor v_r/14 \rfloor + 1) \\
&\quad + 9\lfloor 4v_r/9 \rfloor (\lfloor v_r/14 \rfloor + 1) \cdot (\lfloor 2v_r/9 \rfloor + 1) \ \text{ for } v_r > \frac{9}{2} \\
s_{f,r} &= 5\lfloor 16v_r/9 \rfloor (\lfloor v_r/14 \rfloor + 1) \ \text{ for } v_r < \frac{9}{2} \\
s_{a,r} &= 5(\lfloor v_r/14 \rfloor + 1) + 9(\lfloor v_r/14 \rfloor + 1) \cdot (\lfloor 2v_r/9 \rfloor + 1) \\
&\quad + 5\lfloor 16v_r/9 \rfloor (\lfloor v_r/14 \rfloor + 1) \\
&\quad + 9\lfloor 4v_r/9 \rfloor (\lfloor v_r/14 \rfloor + 1) \cdot (\lfloor 2v_r/9 \rfloor + 1) \ \text{ for } v_r > \frac{9}{2} \\
s_{a,r} &= 14 + \lceil \frac{9 - 2v_r}{\frac{2}{3} v_r - 1} \rceil + 5\lfloor 16v_r/9 \rfloor \ \text{ for } v_r < \frac{9}{2}.
\end{aligned}
$$

We also set $k_{r-1} = s_{a,r} k_r$ for $0 \leq r < s$, and create k_0 copies of M_0. Recall that $k_s = 1$. □

To conclude the lower bound analysis we lower bound the size of s and hence v_s. We observe that for $v_r > \frac{9}{2}$, $s_{a,r} \leq 5(v_r/14 + 1) + 9(v_r/14 + 1) \cdot (2v_r/9 + 1) + 5 \cdot 16v_r/9(v_r/14 + 1) + 4v_r \cdot (v_r/14 + 1) \cdot (2v_r/9 + 1) \leq 4v_r^3/63 + 91v_r^2/63 + 143v_r/9 + 14 \leq 2v_r^3$, as $v_r \geq \frac{9}{2}$, and for $v_r < \frac{9}{2}$, $s_{a,r} \leq 14 + 9 + 80v_r/9 \leq 4v_r^3$.

Note that $v_s = 2\left(\frac{9}{8}\right)^s$. We can conclude that

$$\begin{aligned}
k_0 &\leq 4v_s^3 \cdot 4\left(\frac{8}{9}v_s\right)^3 \ldots 4\left(\left(\frac{8}{9}\right)^{s-1} v_s\right)^3 \\
&= 4\left(2\left(\frac{9}{8}\right)^s\right)^3 \cdot 4\left(2\left(\frac{9}{8}\right)^{s-1}\right)^3 \ldots \cdot \left(2\left(\frac{9}{8}\right)\right)^3 \\
&= (32)^s\left(\frac{9}{8}\right)^{s(s-1)/2} \\
&\leq \left(\frac{9}{8}\right)^{(s^2+58s)/2}.
\end{aligned}$$

Thus the size of item A_0 is at most $17000\left(\frac{9}{8}\right)^{(s^2+58s)/2}$. But this is more than $n/2$ by inspection of the construction. Therefore we can choose s in our construction to satisfy $17000\left(\frac{9}{8}\right)^{(s^2+58s)/2} \geq n/2$; thus

$$s^2 + 58s)/2 \cdot \log\frac{9}{8} + \log 17000 \geq \log n - 1.$$

and hence

$$(s+29)^2 - 29^2 \geq 2(\log n - 1 - \log 17000)/\log\frac{9}{8}$$

So,

$$s \geq (2[(\log n) - 1 - \log 17000 + 142]/\log\frac{9}{8})^{1/2} - 29.$$

This implies

$$\begin{aligned}
v_s &\geq \left(\frac{9}{8}\right)^{(2\log n/\log\frac{9}{8})^{1/2}-29} \\
&\geq 2^{\sqrt{2\log\frac{9}{8}\log n}-29\log\frac{9}{8}} \\
&\geq 2^{\sqrt{\log n/2}-5}.
\end{aligned}$$

APPENDIX C

MEASURING ACCESS TO HEALTH INFORMATION IN DATA-SPARSE REGIONS

C.1 Additional Details on Data and Methodology

Data Cleaning. After generating the initial sets of all queries containing the disease terms ("HIV" or "AIDS," "malaria," or "tuberculosis" or "TB," respectively), removing queries with two or fewer words, and scrubbing the data to remove personal information, including all HIPAA identifiers (CDC, 2003), we manually examined a sample of 1,000 queries from each data set to check whether they were relevant to the disease. The scrubber we used, Tee anonymizer, extracts and replaces PII with more suitable specific placeholders. For example, an email address gets replaced by the text emailpii. There are 13 different types of PII that are replaced including Name, Phone, Address, SSN, CC, and so on. Of these samples, we found that all queries in the malaria data set were related to malaria, but 4.5% of the queries in the HIV/AIDS data set and 16.7% of the queries in the tuberculosis data set were off-topic, containing phrases such as "tb dresses," "TB Joshua," or "4 TB." To improve the quality of the tuberculosis data set, we used these off-topic queries to generate a list of common phrases that we then employed to filter out irrelevant queries from the full tuberculosis data set. After this filtering step, we sampled a fresh set of 1,000 queries and found that only 7.0% were off-topic.

Languages Used. One potential source of bias in our data is that the way in which the data were filtered may have caused us to miss relevant queries made in languages other than English. To understand the extent of this potential problem, we examined the set of all search queries made on Bing anywhere in Africa during February 5–11, 2016. We sampled 1,000 random queries from this set and manually determined how many of these queries appeared to be in English. Overall, 22.5% of the queries were in languages other than English. Of the remaining 77.5%, all were either in English or could potentially have been in English (i.e., the name of a celebrity or the name of a country). Most non-English queries were in either Arabic, French, or Portuguese, and most non-English searches were concentrated in a few countries. Morocco, Algeria, and Egypt had especially high concentrations of non-English searchers; 49% of queries from Morocco, 55% of queries from Algeria, and 61% of queries from Egypt were not in English.

User Demographic Distributions. As reported in the main text, a portion of the queries in our data sets were accompanied by self-reported age and/or gender of the user. Users are identified by an anonymous ID. In this section, we present some analyses to indicate the distribution of these values and associations with population demographic statistics.

We first took all anonymous IDs that are associated with a query in any one of the HIV/AIDS, malaria, or tuberculosis data sets. We look at all users who reported ages between 18 and 50. Recall that this age range corresponds to that considered for our analysis for age and topic usage. (We did not consider individuals who report an age below 18 due to ethical considerations and above 50 due to data sparsity concerns.) We found that this age range account for 90%, 86%, and 87% of the queries made by users with age available for the HIV/AIDS, malaria, and tuberculosis data sets, respectively.

We further analyzed association of the reported age and gender of the users with population statistics. Specifically, we use the statistics presented by DESA (2017), which contains estimated population statistics for different age groups given in 5 year age groups. We took the age ranges between 20 and 49 and consider the fraction of the population in the age ranges 20–24, 25–29, . . . , 45–49. Likewise, we consider the fraction of users who reported ages in these ranges. We find a correlation coefficient of $\rho = 0.9564$ $[0.6478, 0.9954]$ with $p < 0.01$. Note, however, that this correlation coefficient drops significantly when adding older ages due to the sparsity concerns present in our data.

Furthermore, 54.19% percent of anonymous IDs which report their gender are men while 45.81% were women. On the other hand, 50.13 % of the African population is female while 49.87% is male (DESA, 2017). This digital gap observed by gender is consistent with findings that women (as well as older age-groups, families in rural regions, low-literacy individuals, and other underserved communities) have less access to Internet (ITU, 2017). Note, however, accurate statistics disaggregated by the above groups is challenging to obtain in many African nations.

Symptoms (0.93%)	pregnancy, effects, symptoms, early, pathophysiology, treatment, management, effect, pdf, complications, mouth, disease, sign, sore, bitter, throat, nigeria, symptom, dar	malaria lip sores malaria blisters on lips bitterness in mouth and malaria
Natural Cure (1.02%)	cure, natural, home, treat, treatment, remedy, remedies, fever, typhoid, treating, herbal, herbs, medicine, leaf, leaves, good, cures, naturally, lemon	pawpaw leave malaria remedy papaya leaf malaria lipton tea for malaria
Epidemiology (17.39%)	disease, people, year, africa, deaths, download, die, communicable, nigeria, song, number, virus, cases, died, million, caused, mp3, tropical, soty	malaria free sri lanka lyrics malaria theme song stoy- malaria mp3
Drugs (1.31%)	prophylaxis, treatment, quinine, dosage, pregnancy, dose, cdc, doxycycline, prevention, artesunate, children, malarone, chloroquine, severe, table, treat, guidelines, fansidar, treating	fansidar malaria dose quinine maximum dose malaria artefan malaria dose
Breastfeeding (1.05%)	drug, drugs, anti, baby, treat, mother, treatment, breastfeeding, medicine, pregnancy, cancer, fight, good, child, taking, months, affect, medication, babies	malaria breast milk can a breast feeding mother take malaria drugs
Diagnosis (1.24%)	parasite, blood, test, parasites, film, smear, thick, stain, slide, thin, microscope, giemsa, procedure, images, staining, field, medicine, count, density	dar es salaam malaria malaria swamp swollen lip and malaria

Table C.1: Sample LDA topics for the malaria data set with representative words and sample queries.

Symptoms (1.80%)	symptoms, signs, early, stages, warning, list, sign, infection, pulmonary, symtoms, babies, children, infants, symptons, kids, toddlers, cough, baby, symtomps	night sweat in tuberculosis tuberculosis dry cough tb feet and face swelling
Natural Cure (0.85%)	cure, treat, home, treatment, natural, history, remedies, medicine, mdr, group, patient, taboola, utm_source, disease, long, rememdy, traditional, utm_campaign, treatments, herbs	tuberculosis cure discovered does moringa seed cure tb tb reducing natural remedies
Epidemiology (0.93%)	africa, south, statistics, deaths, 2010, death, provinces, sa, stats, show, rate, prevalence, province, 2016, incidence, african, graph, 2015, showing	tb death toll sa tuberculosis graphs tb death provincial statistic
Drug Side-Effects (1.44%)	drugs, effects, side, treatment, anti, medication, effect, drug, line, medications, liver, list, anti-tb, pregnancy, dosage, anti-tuberculosis, patients, adverse, induced	anti-tuberculosis drugs anti tuberculosis combination 2nd line anti tb
Diagnosis (1.28%)	diagnosis, culture, sputum, mycobacterium, gene, test, laboratory, genexpert, smear, xpert, testing, lab, microscopy, expert, negative, stain, diagnostic, procedure, pulmonary, collection	tb auramine staining tb culture sensitivity mycobacterium tuberculosis acid-fast stain
Drug Resistance (1.11%)	drug, resistant, resistance, multidrug, treatment, multi, management, therapy, drugs, multiple, pdf, multi-drug, mycobacterium, patients, antibiotic, mdr, active, rifampicin, latent	multdrug resistant tuberculosis extensively drug resistant vs multidrug resistant tuberculosis

Table C.2: Sample LDA topics for the tuberculosis data set with representative words and sample queries.

Coverage of Africa. Our search data covers all 54 nations of Africa. Figure C.1 shows a heat map of the total search traffic during January 2016–June 2017 period by country and a heat map of the 2016 population of each country. The Spearman correlation coefficient between total search traffic during the January

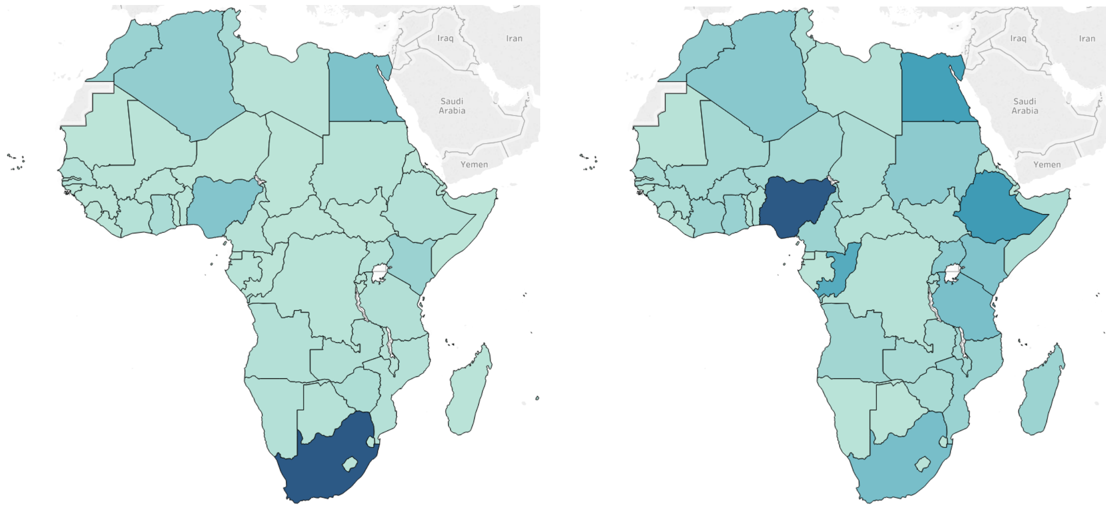


Figure C.1: Left: Heat map of the total search traffic in each country during the January 2016–June 2018 period. Right: Heat map of the population in each country in 2016.

2016–June 2017 period by country and the 2016 population of each country is $\rho = 0.622$ $[0.591 - 0.651]$ with $p < 0.01$. The correlation between search traffic and Internet penetration is $\rho = 0.574$ $[0.540, 0.606]$ with $p < 0.01$.

Building on these insights, we run a multiple linear regression with the HIV prevalence rate as the dependent variable and the fraction of searches associated with the disease, Internet penetration, percent of population in urban settings, population, and GDP as the explanatory variables. We find that the fraction of searches containing the disease name explain some of the variance in disease prevalence even after controlling for the other explanatory variables.

C.2 Analyses of the Popularity of Topics

Topic Popularity by Country. In the main section, we discuss the association between the popularity of the *Stigma* topic for HIV/AIDS in a country and that country's disease prevalence. We ran similar tests on the six topics included in the table in the main body for the malaria and tuberculosis data sets. We find a significant (multi-test corrected) relationship between the popularity of the *Epidemiology* topic for tuberculosis and the tuberculosis incidence rate ($r = 0.509$ multi-test corrected $p < 0.01$). See Figure C.2.[1]

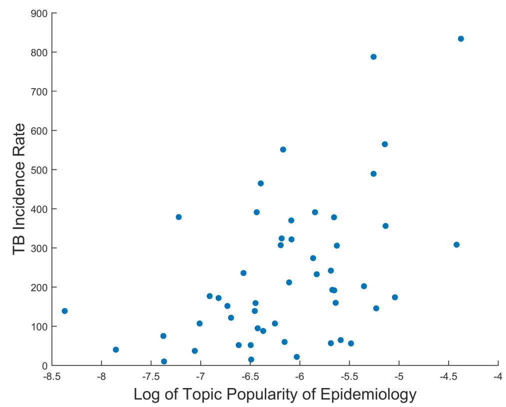


Figure C.2: Popularity of the *Epidemiology* topic versus tuberculosis incidence.

Topic Popularity by User Demographics. As we did for HIV/AIDS, we looked at the topic popularity of the six topics of interest from the topic table in the main section. Tables C.3 and C.4 list the correlation coefficients, where $**$ indicates a p-value of less than 0.01 and $*$ indicates a p-value of less than 0.05. As we saw in the HIV/AIDS data set, women and individuals in the 25–34 age group expressed relatively more interest in topics related to pregnancy, breastfeeding, and family care compared to men in the malaria data set. Additionally, for both the malaria and tuberculosis data sets, users in the 25–34 age group were relatively more interested in symptom-related topics. This is consistent with the literature that nearly half of all new HIV infections occur among the 15–24 age group. The 25–34 age group corresponds to the time-frame where these

[1]As in the main text, relationships are considered significant if they pass the Benjamini-Hochberg false discovery rate (multi-test correction) with $\alpha = 0.5$.

infections may progress significantly, and even develop to AIDS.

	Women	Ages 18–24	Ages 25–34	Ages 35–49
Drug	0.006	-0.068**	-0.008	0.031
Natural Cure	0.027	-0.051**	-0.014	0.051**
Breastfeeding	0.073**	-0.089**	0.060**	0.031
Epidemiology	-0.084**	-0.002	-0.042*	-0.005
Diagnosis	-0.065**	0.032	0.071**	-0.058**
Symptoms	0.055**	0.001	0.062**	-0.019

Table C.3: Relative topic popularity by user demographics for malaria.

	Women	Ages 18–24	Ages 25–34	Ages 35–49
Epidmeology	0.006	0.037*	-0.078**	0.007
Drug Resistance	0.003	0.007	-0.027	0.002
Diagnosis	-0.043*	0.004*	-0.011*	0.001
Symptoms	0.090**	0.014	0.074**	-0.042
Side-effects	0.027	0.028	-0.004	-0.016
Natural Cure	0.018	-0.020	-0.016	0.022

Table C.4: Relative topic popularity by user demographics for TB.

We looked at the topic popularity of the six topics of interest from the topic table in the main section. Women and individuals in the 25–34 age group expressed relatively more interest in topics related to pregnancy, breastfeeding, and family care compared to men in the malaria data set. Additionally, for both the malaria and tuberculosis data sets, users in the 25–34 age group were relatively more interested in symptom-related topics.

C.3 Additional Details on User Behavior and Quality of Results

C.3.1 User Behavior

We examined whether user behavior varies across different topics for the HIV/AIDS, malaria, and tuberculosis data sets. We used four popular metrics in the information retrieval literature: dwell time, maximum dwell time, click count, and successful click count. Dwell time and click count are discussed in the main text. *Maximum dwell time* measures the maximum amount of time that a user spends on any link that is followed. *Successful click count* is the total number of links the user clicks that have a dwell time of at least 30 seconds. We use the same methodology described in the main text and the regression coefficients to report the average values. Results are shown in Figure C.3. Note that for the HIV/AIDS and malaria data sets, users issuing queries associated with *Natural cure* exhibited relatively low activity (by all four metrics) compared to many of the other topics of interest; this is not true for the tuberculosis data set.

C.3.2 Quality of Content

To measure the quality of the links presented by topic, we examined the set of links returned in the first position for the thirty most representative queries for each of the six topics from the malaria and tuberculosis data sets that appear in the table in the main text, in the same way described in the main text for the HIV/AIDS data set. Each link was evaluated by three research assistants, each

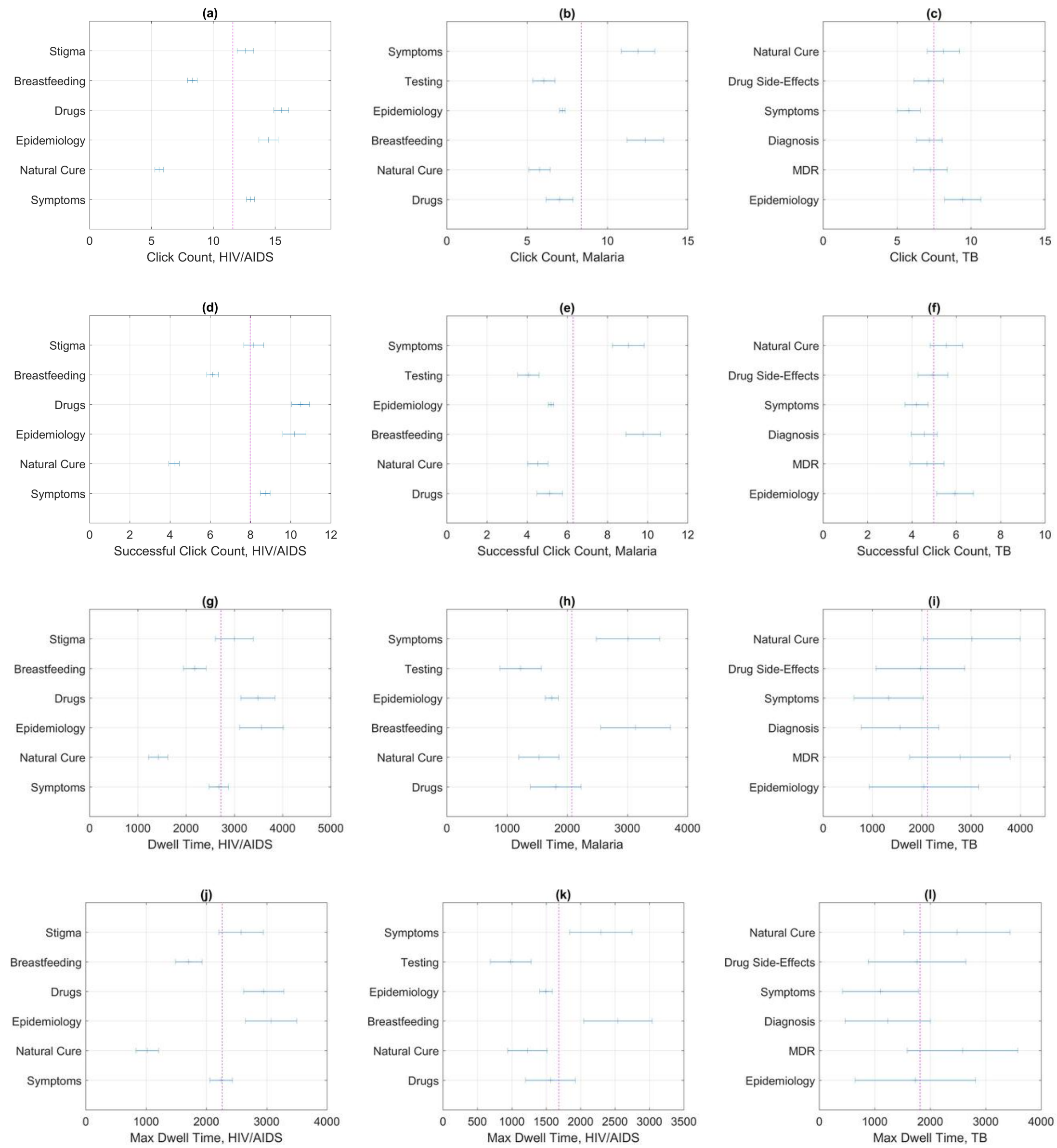


Figure C.3: Rows, from top to bottom: average click count, successful click count, dwell time, and maximum dwell time for queries associated with different topics. Columns, from left to right: HIV/AIDS, malaria, and tuberculosis data sets. The vertical lines represent the mean values across topic.

of whom has graduate-level training in medicine or public health, and at least one of whom specializes in the corresponding disease. In all three cases, the research assistants were asked to assess the relevance, accuracy, and objectiveness of the links returned. In particular, the research assistants were presented with the following questions.

- *Relevance:* How comprehensive and complete is the information provided on the website and does it appear to provide the right level of detail? Is the URL related to the disease?
- *Accuracy:* Are sources of information properly identified, and are there any glaring omissions or misinformation?
- *Objectiveness:* Does the information presented appear in an objective manner without political, cultural, religious, or institutional bias?

The research assistants were asked to assign a single rating for each link on a scale from 1 to 5, with 1 equal to bad quality and 5 equal to high quality. Values were defined as:

1. Bad quality: several serious issues concerning all three of relevance, accuracy, and objectiveness
2. Subpar quality: several serious issues covering at least two of relevance, accuracy, or objectiveness
3. Mediocre quality: several issues concerning relevance, accuracy, or objectiveness
4. Good quality: mostly relevant, accurate, and objective, with a few small issues
5. High quality: very relevant, accurate, and objective

Note that the research assistants were not asked to take into account the website design, interface, usability, or other metrics unrelated to content.

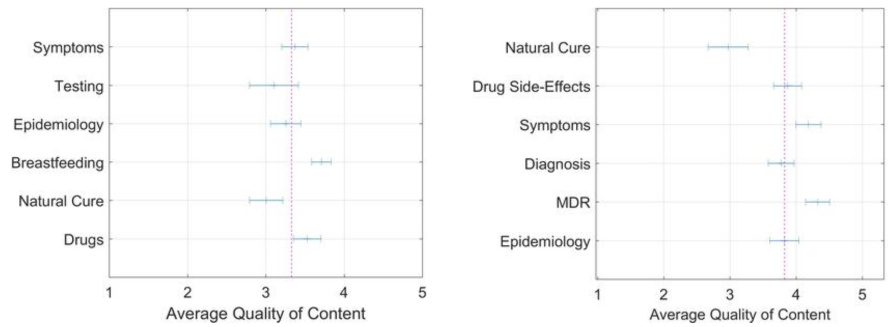


Figure C.4: Average quality of content of webpages returned to users for the 30 queries most strongly associated with the six malaria (left) and tuberculosis (right) topics.

Consistent with the observations for the HIV/AIDS data set, for the tuberculosis data set, the quality of content returned to users was, on average, rated lower for queries related to the *Natural cure* topic than for other topics of interest. In contrast, for the malaria data set, the quality of links returned was indistinguishable among queries related to the *Natural cure*, *Epidemiology*, *Testing*, and *Symptoms* topics.

APPENDIX D

FACILITATING NETWORK INTEGRATION

We present missing proofs and additional discussions in this appendix.

D.1 Triadic Closure and Stochastic Block Models

Recall that for ease of presentation in the main text we presented results for the case where $k = 2$. We remove this assumption and present generalized results below.

Lemma 45. Given a network G resulting from an SBM with $k \geq 2$, the expected number of monochromatic wedges is:

$$\sum_{i=1}^{k} \left(3\binom{n/k}{3} p^2(1-p) + (n - n/k)\binom{n/k}{2} q^2(1-p) \right) \tag{D.1}$$

while that of bichromatic wedges is

$$\sum_{i=1}^{k} 2(n - n/k)\binom{n/k}{2} pq(1-q). \tag{D.2}$$

Proof. The first term in Equation D.1 results from the expected number of monochromatic wedges from a monochromatic triplets is where $\binom{n/k}{3}$ counts the number of monochromatic triplets of each type, 3 is the number of ways to choose the center of the wedge, and $p^2(1-p)$ counts the likelihood that such a triplet results in a wedge. The second term results from the expected number of monochromatic wedges from a bichromatic triplet, where $(n-n/k)\binom{n/k}{2}$ counts the number of bichromatic triplets and $q^2(1-p)$ corresponds to the likelihood that such a triplet results in a monochromatic wedge.

Likewise, for Equation D.2, we note that $(n - n/k)\binom{n/k}{2}$ counts the number bichromatic triplets. Such a triplet results in a bichromatic wedge with probability $pq(1-q)$. $\square$

We can similarly analyze the effect of triadic closure on network integration for $k \geq 2$.

Theorem 47. Given an SBM with $k \geq 2$, there is a sufficiently large n such that triadic closure increases network integration if $p > q$.

Proof. The fraction of bichromatic edges, $f(G)$, in this general case is:

$$\begin{aligned}
f(G) &= \frac{\frac{1}{2}\sum_i \frac{n}{k}\left(n - \frac{n}{k}\right)q}{\frac{1}{2}\sum_i \frac{n}{k}\left(n - \frac{n}{k}\right)q + \sum_i \binom{n/k}{2}p} \\
&= \frac{\frac{n}{2}\left(n - \frac{n}{k}\right)q}{\frac{n}{2}\left(n - \frac{n}{k}\right)q + \frac{k}{2}\left(\frac{n}{k}\right)^2 p} \\
&= \frac{\left(\frac{n^2}{2} - \frac{n^2}{2k}\right)q}{\left(\frac{n^2}{2} - \frac{n^2}{2k}\right)q + \frac{n^2}{2k}p} + o(1) \\
&= \frac{q(k-1)}{q(k-1) + p} + o(1)
\end{aligned}$$

Using the above lemma, we can simplify the fraction of bichromatic wedges to be

$$w(G) = \frac{(k-1)pq(1-q)}{(k-1)pq(1-q) + \frac{p^2(1-p)}{2} + \frac{k-1}{2}q^2(1-p)} + o(1)$$

As above, we are interested in the inequality $f(G) \leq w(G)$ for large enough n.

That is,

$$
\begin{aligned}
&\frac{q(k-1)}{q(k-1)+p} \leq \frac{(k-1)pq(1-q)}{(k-1)pq(1-q)+\frac{p^2(1-p)}{2}+\frac{k-1}{2}q^2(1-p)} \\
&\frac{(k-1)^3q^3(1-p)}{2}+\frac{(k-1)qp^2(1-p)}{2} \leq (k-1)qp^2(1-q) \\
&(k-1)^2q^2(1-p)+p^2(1-p) \leq 2p^2(1-q) \\
&(k-1)^2q^2(1-p) \leq p^2(1-2q+p)
\end{aligned}
$$

Note that when $k = 2$, this last inequality simplifies to that given in the main text.

We therefore look at the expression

$$p^2(1-2q+p)-(k-1)^2q^2(1-p).$$

For $k > 2$ and $q \in [0, 1]$, this expression as roots at

$$q = \frac{p^2 - \sqrt{-kp^4+2p^4+kp^2-p^2}}{(p-1)(k-1)}.$$

Note that this value also simplifies to that given in the main text for $k = 2$. Following the same argument as in the text, we therefore have that triadic closure preserves network integration for $q = \frac{p^2-\sqrt{-kp^4+2p^4+kp^2-p^2}}{kp-k-p+1}$ and increases integration for q less than this value.

The expression

$$p^2(1-2q+p)-(k-1)^2q^2(1-p)$$

simplifies to $p^2(1-p)k(2-k)$ for $q = p$. This value is positive if $k > 2$. Therefore, the insight that triadic closure increases network integration in the setting where nodes exhibit type-bias continues to hold for $k \geq 2$.

□

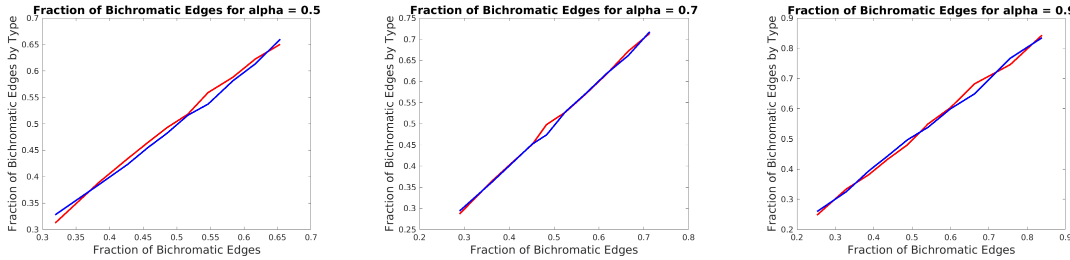


Figure D.1: Values for $f_R(t)$, $f_B(t)$, and $f(t)$ as N_S goes from 1 to 10 for $N_D = 10 - N_S$ and $N_F = 20$.

D.2 A Dynamic Network Formation Model

We present a heuristic argument to support Identity 1 in this section. First, we compare $f_R(t)$, $f_B(t)$, and $f(t)$. By running numerous simulations on various combinations parameters, we find that $f_R(t)$ and $f_B(t)$ both also converge to f^* for large enough t. This observation agrees with the intuition that the model proceeds by adding a node of a random type then proceeding symmetrically regardless of the node type. Therefore, as the network approaches the equilibrium state, we can also expect the $f_R(t)$ and $f_B(t)$ to approach this value by symmetry.

Figure D.1 presents a sample of our results. We run the plots on $n = 500$ nodes and average over 10 trials. We note that these values on the fraction of bichromatic edges, including those disaggregated by type, converge to the expected values relatively quickly. And, most notably, $f_R(t)$ and $f_B(t)$ show comparable values for a range of α values. Note that we run these plots for a combination of $N_S + N_D$ and N_F values and the qualitative insights remain the same. Furthermore, the red and blue curves converge to the point of indistinguishableness as we increase n and the number of trials, further supporting our argument.

Lemma 55. *The fraction of bichromatic edges of $G(t)$ as t goes to infinity converges to:*

$$f(t) = \frac{N_D + (1-\alpha)N_F}{N_D + N_S + 2(1-\alpha)N_F}.$$

Proof. In equilibrium, $f(t+1) = f(t)$. Therefore:

$$\begin{aligned}
\frac{b(t)}{m(t)} &= \frac{b(t) + N_D + \alpha N_F f(t) + (1-\alpha)N_F(1-f(t))}{m(t)+N} \\
Nb(t) &= m(t)\left(N_D + \alpha N_F f(t) + (1-\alpha)N_F(1-f(t))\right) \\
Nf(t) &= N_D + \alpha N_F f(t) + (1-\alpha)N_F - (1-\alpha)N_F f(t) \\
Nf(t) &= N_D + (2\alpha - 1)N_F f(t) + (1-\alpha)N_F \\
f(t) &= \frac{N_D + (1-\alpha)N_F}{N + (1-2\alpha)N_F}
\end{aligned}$$

This last equality holds since $N = N_S + N_D + N_F$.

□

D.3 Interventions to Alleviate Network Segregation

Lemma 50. Suppose we are given a network $G(0)$, values N_F, N, and α, and a time-horizon $[t_1, t_2]$. Starting from $f(t_1)$, the intervention which maximizes $f(t_2)$ is that which sets $N_D = N - N_F$ over the entire time-horizon $[t_1, t_2]$.

Proof. We first construct the feasible region shown in Figure 9.2. Recall that the green lines correspond to the intervention which sets $N_D = N - N_F$ and the red lines correspond to the intervention which set $N_D = 0$. The thick lines for green and red correspond to the interventions that use this intervention starting from $[t_1, f_1]$.

Take the path resulting from setting $N_D = N - N_F$ for each time step t for $t \in [t_1, t_2]$ by ℓ and denote the resulting $f(t_2)$ by $f_\ell(t_2)$. We want to prove that there exists no intervention ℓ' that sets $N_{D'} < N - N_F$ at any time step t such that the resulting$f_{\ell'}(t_2) > f_\ell(t_2)$. That is, the path ℓ' lies below ℓ. By contradiction, suppose that such a path ℓ' exists. Find value t_i such that for all $t_1 < t < t_i$, $f_\ell(t_i) \geq f_{\ell'}(t_i)$ but $f_\ell(t_{i+1}) < f_{\ell'}(t_{i+1})$. This contradicts the observation that a strategy that locally maximizes network integration is that which sets $N_D = N - N_F$. □

BIBLIOGRAPHY

GiveDirectly: Send money directly to the extreme poor. https://givedirectly.org/.

Public law 104-208, September 1996. URL https://www.govinfo.gov/content/pkg/PLAW-104publ208/pdf/PLAW-104publ208.pdf.

A world that counts: Mobilising the data revolution for sustainable development. Technical report, November 2014.

The Eviction Lab. https://evictionlab.org/, 2018.

The World Bank DataBank: World Development Index Database Archives, 2018.

Abel Wudenah Abamagal and Philipos Wudineh Abamagal. Role and determinants of saving and credit cooperative on financial intermediation in hadiya zone, southern ethiopia. *Economics*, 8(4):134–141, 2019.

Atila Abdulkadiroğlu and Tayfun Sönmez. Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica*, 66(3):689–701, 1998.

Atila Abdulkadiroğlu and Tayfun Sönmez. School choice: A mechanism design approach. *American economic review*, 93(3):729–747, 2003.

Atila Abdulkadirolu and Tayfun Sönmez. Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica*, 66:689–701, May 1998.

Rediet Abebe and Kira Goldner. MD4SG research group. http://md4sg.com/researchgroup/.

Rediet Abebe and Kira Goldner. Mechanism design for social good. *AI Matters*, 4(3):27–34, 2018a.

Rediet Abebe and Kira Goldner. A report on the workshop on mechanism design for social good. *ACM SIGecom Exchanges*, 16(2):2–11, 2018b.

Rediet Abebe, Salvatore Giorgi, Anna Tedijanto, Anneke Buffone, and H. Andrew Schwartz. Quantifying community characteristics of maternal mortality using social media, a. Preprint.

Rediet Abebe, Irene Lo, and Ana-Andreea Stoica. MD4SG Initiative. `http://md4sg.com/`, b.

Rediet Abebe, Jon Kleinberg, and David C. Parkes. Fair division via social comparison. In *Proc. of the 16th International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '17, 2017.

Rediet Abebe, Lada A Adamic, and Jon Kleinberg. Mitigating overexposure in viral marketing. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

Rediet Abebe, Richard Cole, Vasilis Gkatzelis, and Jason D Hartline. A truthful cardinal mechanism for one-sided matching. *arXiv preprint arXiv:1903.07797*, 2019a.

Rediet Abebe, Shawndra Hill, Jennifer Wortman Vaughan, Peter M Small, and H Andrew Schwartz. Using search queries to understand health information needs in Africa. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, pages 3–14, 2019b.

Rediet Abebe, Nicole Immorlica, Jon Kleinberg, and Brendan Lucier. Triadic closure alleviates network integration, 2019c.

Rediet Abebe, Solon Barocas, Jon Kleinberg, Karen Levy, Manish Raghavan, and David Robinson. Roles for computing in social change. In *Conference on Fairness, Accountability and Transparency*, 2020a.

Rediet Abebe, Jon Kleinberg, and S. Matthew Weinberg. Subsidy allocations in the presence of income shocks. In *Thirty-Fourth AAAI Conference on Artificial Intelligence*, 2020b.

Hal Abelson and et. al. Letter to the Honorable Elaine C. Duke. `https://www.brennancenter.org/sites/default/files/Technology\%20Experts\%20Letter\%20to\%20DHS\%20Opposing\%20the\%20Extreme\%20Vetting\%20Initiative\%20-\%2011.15.17.pdf`, 2016.

Harold Abelson, Ross Anderson, Steven M Bellovin, Josh Benaloh, Matt Blaze, Whitfield Diffie, John Gilmore, Matthew Green, Susan Landau, Peter G Neumann, et al. Keys under doormats: mandating insecurity by requiring government access to all data and communications. *Journal of Cybersecurity*, 1(1): 69–79, 2015.

Julie Abimanyi-Ochom, Hasheem Mannan, Nora Ellen Groce, and Joanne McVeigh. HIV/AIDS knowledge, attitudes and behaviour of persons with and without disabilities from the Uganda Demographic and Health Survey 2011: Differential access to HIV/AIDS information and services. *PLOS one*, 2017.

Lada A. Adamic and Natalie Glance. The political blogosphere and the 2004 U.S. election: Divided they blog, 2005.

Philip E Agre. Lessons learned in trying to reform AI. *Social science, technical systems, and cooperative work: Beyond the Great Divide*, page 131, 1997.

AHO. African Health Observatory, 2010.

S Rao Aiyagari. Uninsured idiosyncratic risk and aggregate saving. *The Quarterly Journal of Economics*, 109(3):659–684, 1994.

Jon Aizen, Dan Huttenlocher, Jon Kleinberg, and Antal Novak. Traffic-based feedback on the Web. 101(Suppl.1):5254–5260, 2004.

George A Akerlof. The economics of "tagging" as applied to the optimal income tax, welfare programs, and manpower planning. *The American economic review*, 68(1):8–19, 1978.

Saeed Alaei, Hu Fu, Nima Haghpanah, Jason Hartline, and Azarakhsh Malekian. Bayesian optimal auctions via multi-to single-agent reduction. *arXiv preprint arXiv:1203.5099*, 2012.

Saeed Alaei, Pooya Jalaly Khalilabadi, and Eva Tardos. Computing equilibrium in matching markets. In *ACM Conference on Economics and Computation, EC '17*, pages 245–261. ACM, 2017.

Vivi Alatas, Ririn Purnamasari, Matthew Wai-Poi, Abhijit Banerjee, Benjamin A Olken, and Rema Hanna. Self-targeting: Evidence from a field experiment in indonesia. *Journal of Political Economy*, 124(2):371–427, 2016.

Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. Discrimination through optimization: How Facebook's ad delivery can lead to skewed outcomes. *arXiv preprint arXiv:1904.02095*, 2019.

Sabina Alkire and James Foster. Counting and multidimensional poverty measurement. *Journal of public economics*, 95(7):476–487, 2011.

Winston J Allen, Albert Ouedraogo, and Leela McCullough. Health information needs in West Africa: Results of a survey on the role of the West Africa Health Organization (WAHO), 2010.

Kristen M Altenburger and Johan Ugander. Monophily in social networks introduces similarity among friends-of-friends. *Nature human behaviour*, 2(4):284, 2018.

Benjamin M Althouse, Yih Yng Ng, and Derek AT Cummings. Prediction of dengue incidence using search query surveillance. *PLoS neglected tropical diseases*, 2011.

Brent W Ambrose and Moussa Diop. Information asymmetry, regulations, and equilibrium outcomes: Theory and evidence from the housing rental market. 2016.

Sudhir Anand and Amartya Sen. Concepts or human development and poverty! a multidimensional perspective. *United Nations Development Programme, Poverty and human development: Human development papers*, pages 1–20, 1997.

Sinan Aral and Dylan Walker. Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management science*, 57(9):1623–1639, 2011.

Nick Arnosti and Peng Shi. Design of lotteries and waitlists for affordable housing allocation. 2018.

David Arthur, Rajeev Motwani, Aneesh Sharma, and Ying Xu. Pricing strategies for viral marketing on social networks. 2009.

Søren Asmussen and Hansjörg Albrecher. *Ruin probabilities*. World Scientific Publishing Co Pte Ltd, 2010.

Esso-Hanam Atake. Health shocks in sub-Saharan Africa: are the poor and uninsured households more vulnerable? *Health economics review*, 8(1):26, 2018.

Anthony B Atkinson. Multidimensional deprivation: contrasting social welfare and counting approaches. *The Journal of Economic Inequality*, 1(1):51–65, 2003.

Anthony B Atkinson. *Inequality*. Harvard University Press, 2015.

Ellen Averett, John S Neuberger, Gail Hansen, and Michael H Fox. Evaluation of West Nile virus education campaign. *Emerging infectious diseases*, 11(11): 1751, 2005.

Chen Avin, Barbara Keller, Zvi Lotker, Claire Mathieu, David Peleg, and Yvonne-Anne Pignolet. Homophily and the glass ceiling effect in social networks, 2015.

Pablo Azcue and Nora Muler. Optimal investment strategy to minimize the ruin probability of an insurance company under borrowing constraints. *Insurance: Mathematics and Economics*, 44(1):26–34, 2009.

Haris Aziz, Aris Filos-Ratsikas, Jiashu Chen, Simon Mackenzie, and Nicholas Mattei. Egalitarianism of random assignment mechanisms: (extended abstract). In *International Conference on Autonomous Agents and Multiagent Systems*, pages 1267–1268, 2016.

Patrick Bajari and Ali Hortascu. Economic insights from Internet auctions. 42: 457–486, 2004.

Jack M Balkin. Digital speech and democratic culture: A theory of freedom of expression for the information society. *NYU Law Review*, 79:1, 2004.

Abhijit Banerjee, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson. The diffusion of microfinance. *Science*, 341(6144), 2013.

Chelsea Barabas, Karthik Dinakar, Joichi Ito, Madars Virza, and Jonathan Zittrain. Interventions over predictions: Reframing the ethical debate for actuarial risk assessment. *arXiv preprint arXiv:1712.08238*, 2017.

Moshe Barach and John J Horton. How do employers use compensation history?: Evidence from a field experiment. 2017.

Siddharth Barman, Sanath Kumar Krishnamurthy, and Rohit Vaish. Finding fair and efficient allocations. In *ACM Conference on Economics and Computation, EC '18*, pages 557–574, 2018.

Solon Barocas and Andrew D Selbst. Big data's disparate impact. *California Law Review*, 104:671, 2016.

Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*, 1995.

Cynthia L Bennett and Os Keyes. What is the point of fairness? Disability, AI and the complexity of justice. *arXiv preprint arXiv:1908.01024*, 2019.

Sebastian Benthall. Computational institutions. `https://digifesto.com/2018/12/22/computational-institutions/`, Dec 2018a.

Sebastian Benthall. The politics of AI ethics is a seductive diversion from fixing our broken capitalist system, Dec 2018b.

Jonah Berger and Chip Heath. Where consumers diverge from others: Identity signaling and product domains. *Journal of Consumer Research*, 34(2):121–134, 2007.

Marianne Bertrand and Sendhil Mullainathan. Are Emily and Greg more employable than Lakisha and Jamal? a field experiment on labor market discrimination. *American economic review*, 94(4):991–1013, 2004.

Asia J Biega, Krishna P Gummadi, and Gerhard Weikum. Equity of attention: Amortizing individual fairness in rankings. *arXiv preprint arXiv:1805.01788*, 2018.

Martina Björkman-Nyqvist. Income shocks and gender gaps in education: Evidence from Uganda. *Journal of Development Economics*, 105:237–253, 2013.

Christopher Blattman and Paul Niehaus. Show them the money: Why giving cash helps alleviate poverty. *Foreign Affairs*, 93(3):117–126, 2014.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent Dirichlet allocation. *JMLR*, 3, March 2003. ISSN 1532-4435.

Anna Bogomolnaia and Hervé Moulin. A new solution to the random assignment problem. *Journal of Economic Theory*, 100(2):295 – 328, 2001. ISSN 0022-0531.

Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In *Advances in neural information processing systems*, pages 4349–4357, 2016.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

Yann Bramoullé, Sergio Currarini, Matthew O Jackson, Paolo Pin, and Brian W Rogers. Homophily and long-run integration in social networks. *Journal of Economic Theory*, 147(5), 2012.

Steven Brams and Alan Taylor. *Fair Division: From Cake Cutting to Dispute Resolution*. 1996.

Simina Brânzei, Vasilis Gkatzelis, and Ruta Mehta. Nash social welfare approximation for strategic agents. In *ACM Conference on Economics and Computation, EC '17*, pages 611–628, 2017.

Sarah Brayne. Big data surveillance: The case of policing. *American Sociological Review*, 82(5):977–1008, 2017.

Meredith Broussard. *Artificial Unintelligence: How Computers Misunderstand the World*. 2018.

Meredith Broussard, Nicholas Diakopoulos, Andrea L. Guzman, Rediet Abebe, Michel Dupagne, and Ching-Hua Chuan. Artificial intelligence and journalism. *Journalism & Mass Communication Quarterly*, 96(3):673–695, 2019. doi: 10.1177/1077699019859901. URL `https://doi.org/10.1177/1077699019859901`.

John S Brownstein, Clark C Freifeld, and Lawrence C Madoff. Digital disease detectionharnessing the web for public health surveillance. *New England Journal of Medicine*, 2009.

Waylon Brunette, Mitchell Sundt, Nicola Dell, Rohit Chaudhri, Nathan Breit, and Gaetano Borriello. Open data kit 2.0: expanding and refining information services for developing regions. In *Proceedings of the 14th Workshop on Mobile Computing Systems and Applications*, page 10. ACM, 2013.

Gharad Bryan, Jonathan De Quidt, Tom Wilkening, and Nitin Yadav. Land trade and development: A market design approach. 2017.

Eric Budish. The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes. *Journal of Political Economy*, 119(6):1061–1103, 2011.

Eric Budish, Gérard P Cachon, Judd B Kessler, and Abraham Othman. Course match: A large-scale implementation of approximate competitive equilibrium from equal incomes for combinatorial allocation. *Operations Research*, 65(2): 314–336, 2016.

Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency*, pages 77–91, 2018.

John W. Byers, Michael Mitzenmacher, and Georgios Zervas. The groupon effect on yelp ratings: A root cause analysis. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, EC '12, pages 248–265, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1415-2. doi: 10.1145/2229012.2229034. URL `http://doi.acm.org/10.1145/2229012.2229034`.

Guido Calabresi and Philip Bobbitt. Tragic choices. 1984.

Aylin Caliskan, Joanna J Bryson, and Arvind Narayanan. Semantics derived automatically from language corpora contain human-like biases. *Science*, 356 (6334):183–186, 2017.

Antoni Calvo-Armengol and Matthew O Jackson. The effects of social networks on employment and inequality. *American Economic Review*, 94(3), 2004.

Antoni Calvo-Armengol, Eleonora Patacchini, and Yves Zenou. Peer effects and social networks in education. *The Review of Economic Studies*, 76(4), 2009.

Donald T. Campbell. Assessing the impact of planned social change. *Evaluation and Program Planning*, 2:67–90, 1979.

Ioannis Caragiannis, David Kurokawa, Hervé Moulin, Ariel D. Procaccia, Nisarg Shah, and Junxing Wang. The unreasonable fairness of maximum Nash welfare. In *ACM Conference on Economics and Computation, EC '16*, pages 305–322, 2016.

Herman Anthony Carneiro and Eleftherios Mylonakis. Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clinical infectious diseases*, pages 1557–1564, 2009.

Ruggiero Cavallo. Optimal decision-making with minimal waste: Strategyproof redistribution of VCG payments. In *International Conference on Autonomous Agents and Multiagent Systems*, pages 882–889. ACM, 2006.

CDC. Active bacterial core surveillance report, emerging infections program network, streptococcus pneumoniae, 1998.

CDC. HIPAA privacy rule and public health. guidance from cdc and the us department of health and human services. *MMWR: Morbidity and mortality weekly report*, 52(Suppl. 1):1–17, 2003.

Brian Chan and Alexander Tsai. Trends in HIV-related stigma in the general population during the era of antiretroviral treatment expansion: an analysis of 31 sub-Saharan African countries, 2015.

Brian T Chan and Alexander C Tsai. HIV knowledge trends during an era of rapid antiretroviral therapy scale-up: an analysis of 33 sub-Saharan African countries. *Journal of the International AIDS Society*, 2018.

Emily H Chan, Vikram Sahai, Corrie Conrad, and John S Brownstein. Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. *PLoS neglected tropical diseases*, 2011.

Eric Chan. Information problems in education. 2017.

Yeon-Koo Che and Fuhito Kojima. Asymptotic equivalence of probabilistic serial and random priority mechanisms. *Econometrica*, 78(5):1625–1672, 2010. ISSN 1468-0262.

Angela Chen. Inmates in finland are training AI as part of prison labor. *The Verge*, 2019.

Le Chen, Ruijun Ma, Anikó Hannák, and Christo Wilson. Investigating the impact of gender on rank in resume search engines. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 651. ACM, 2018.

Alexandra Chouldechova. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data*, 5(2):153–163, 2017.

Angèle Christin. Counting clicks: Quantification and variation in web journalism in the united states and france. *American Journal of Sociology*, 123(5): 1382–1415, 2018.

George Christodoulou, Aris Filos-Ratsikas, Søren Kristoffer Stiil Frederiksen, Paul W Goldberg, Jie Zhang, and Jinshan Zhang. Social welfare in one-sided matching mechanisms. In *International Conference on Autonomous Agents and Multiagent Systems*, pages 30–50, 2016.

Danielle Keats Citron. Technological due process. *Washington University Law Review*, 2018.

David A Clarke. Digital data processing arbitration system, December 6 1988. US Patent 4,789,926.

Stephen Coate and Glenn C Loury. Will affirmative-action policies eliminate negative stereotypes? *The American Economic Review*, pages 1220–1240, 1993.

Devin Coldewey. Sen. Harris tells federal agencies to get serious about facial recognition risks. *TechCrunch*, 2018.

Richard Cole and Vasilis Gkatzelis. Approximating the Nash social welfare with indivisible items. *SIAM Journal of Computing*, 47(3):1211–1236, 2018.

Richard Cole, Vasilis Gkatzelis, and Gagan Goel. Mechanism design for fair division: Allocating divisible items without payments. In *Proceedings of the Fourteenth ACM Conference on Electronic Commerce*, EC '13, pages 251–268. ACM, 2013. ISBN 978-1-4503-1962-1.

Josh Compton and Michael Pfau. Spreading inoculation: Inoculation, resistance to influence, and word-of-mouth communication. *Communication Theory*, 19 (1):9–28, 2009.

Concerned Researchers. On recent research auditing commercial facial analysis technology. *Medium.com*, 2019.

Sasha Costanza-Chock. Design justice, ai, and escape from the matrix of domination. *Journal of Design and Science*, 2018.

National Research Council et al. *Toward an Integrated Science of Research on Families: Workshop Report*. National Academies Press, 2011.

Karen Cravens, Elizabeth Goad Oliver, and Sridhar Ramamoorti. The reputation index:: Measuring and managing corporate reputation. *European Management Journal*, 21(2):201 – 212, 2003. ISSN 0263-2373. doi: http://dx.doi.org/10.1016/S0263-2373(03)00015-X. URL //www.sciencedirect.com/science/article/pii/S026323730300015X.

Kate Crawford. You and AI – the politics of AI. https://www.youtube.com/watch?v=HPopJb5aDyA. Accessed: 2019-08-01.

Kate Crawford and Ryan Calo. There is a blind spot in AI research. *Nature News*, 538(7625):311, 2016.

Anca E. Cretu and Roderick J. Brodie. The influence of brand image and company reputation where manufacturers market to small firms: A customer value perspective. *Industrial Marketing Management*, 36(2):230 – 240, 2007. ISSN 0019-8501. doi: http://dx.doi.org/10.1016/j.indmarman.2005.08.013. URL //www.sciencedirect.com/science/article/pii/S0019850105001331. Project Marketing and the Marketing of Solutions.

Sergio Currarini, Matthew O Jackson, and Paolo Pin. An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4):1003–1045, 2009.

Krishna Dasaratha. Distributions of centrality on networks. *arXiv preprint arXiv:1709.10402*, 2017.

Amit Datta, Michael Carl Tschantz, and Anupam Datta. Automated experiments on ad privacy settings. *Proceedings on privacy enhancing technologies*, 2015(1):92–112, 2015.

Aaron C Davis and Wesley Lowery. FBI director calls lack of data on police shootings 'ridiculous,' 'embarrassing'. *The Washington Post*, 7, 2015.

Maria De-Arteaga, William Herlands, Daniel B Neill, and Artur Dubrawski. Machine learning for the developing world. *ACM Transactions on Management Information Systems (TMIS)*, 9(2):9, 2018.

Maria De Bruyn. Gender, adolescents and the HIV/AIDS epidemic: the need for comprehensive sexual and reproductive health responses. 2000.

Munmun De Choudhury and Sushovan De. Mental health discourse on reddit: Self-disclosure, social support, and anonymity., 2014.

Munmun De Choudhury, Meredith Ringel Morris, and Ryen White. Seeking and sharing health information online: Comparing search engines and social media, 2014.

David Delacrétaz, Scott Duke Kominers, and Alexander Teytelboym. Refugee resettlement. *University of Oxford Department of Economics Working Paper*, 2016.

Lina Dencik, Fieke Jansen, and Philippa Metcalfe. A conceptual framework for approaching social justice in an age of datafication. *DATAJUSTICE project*, 30, 2018.

DESA. World Population Prospects: The 2017 Revision, 2017. United Nations: Department of Economic and Social Affairs.

Rishi Desai, Aron J Hall, Benjamin A Lopman, Yair Shimshoni, Marcus Rennick, Niv Efron, Yossi Matias, Manish M Patel, and Umesh D Parashar. Norovirus disease surveillance using Google Internet query share data. *Clinical Infectious Diseases*, pages e75–e78, 2012.

Matthew Desmond. Eviction and the reproduction of urban poverty. *American Journal of Sociology*, 118(1):88–133, 2012.

Matthew Desmond. Unaffordable America: Poverty, housing, and eviction. *Fast Focus: Institute for Research on Poverty*, 22(22):1–6, 2015.

Matthew Desmond. *Evicted: Poverty and profit in the American city*. Broadway Books, 2016.

Matthew Desmond and Carl Gershenson. Housing and employment insecurity among the working poor. *Social Problems*, 63(1):46–67, 2016.

Matthew Desmond and Tracey Shollenberger. Forced displacement from rental housing: Prevalence and neighborhood consequences. *Demography*, 52(5): 1751–1772, 2015.

Nikhil R Devanur and Ravi Kannan. Market equilibria in polynomial time for fixed number of goods or agents. In *IEEE Symposium on Foundations of Computer Science, 2008. FOCS'08*, pages 45–53, 2008.

Catherine D'Ignazio and Lauren Klein. Data feminism, 2019.

Paul DiMaggio and Filiz Garip. How network externalities can exacerbate intergroup inequality. *American Journal of Sociology*, 116(6):1887–1933, 2011.

Taryn Dinkelman, David Lam, and Murray Leibbrandt. Linking poverty and income shocks to risky sexual behaviour: evidence from a panel study of young adults in cape town. *South African Journal of Economics*, 76:S52–S74, 2008.

Pedro Domingos and Matt Richardson. Mining the network value of customers. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '01, pages 57–66, New York,

NY, USA, 2001. ACM. ISBN 1-58113-391-X. doi: 10.1145/502512.502525. URL http://doi.acm.org/10.1145/502512.502525.

Yuxiao Dong, Reid A Johnson, Jian Xu, and Nitesh V Chawla. Structural diversity and homophily: A study across more than one hundred big networks, 2017.

Nathan Eagle, Michael Macy, and Rob Claxton. Network diversity and economic development. *Science*, 328(5981):1029–1031, 2010.

David Easley and Jon Kleinberg. *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, 2010.

Jonathan Eaton and Harvey S Rosen. Labor supply, uncertainty, and efficient taxation. *Journal of Public Economics*, 14(3):365–374, 1980a.

Jonathan Eaton and Harvey S Rosen. Optimal redistributive taxation and uncertainty. *The Quarterly Journal of Economics*, 95(2):357–364, 1980b.

Jonathan Eaton and Harvey S Rosen. Taxation, human capital, and uncertainty. *The American Economic Review*, 70(4):705–715, 1980c.

Laurel Eckhouse, Kristian Lum, Cynthia Conti-Cook, and Julie Ciccolini. Layers of bias: A unified approach for understanding problems with risk assessment. *Criminal Justice and Behavior*, 46(2):185–2019, 2018.

Michael Ekstrand and Karen Levy. FAT* network. https://fatconference.org/network/. Accessed: 2019-08-01.

Michael D Ekstrand and Martijn C Willemsen. Behaviorism is not enough: better recommendations through listening to users. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 221–224. ACM, 2016.

Jon Elster. *Local justice: How institutions allocate scarce goods and necessary burdens*. Russell Sage Foundation, 1992.

Virginia Eubanks. *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press, 2018a.

Virginia Eubanks. Automating bias. *Scientific American*, 319(5):68–71, 2018b.

Gunther Eysenbach and Christian Köhler. How do consumers search for and appraise health information on the world wide web? qualitative study using focus groups, usability tests, and in-depth interviews. *BMJ*, 2002.

Gunther Eysenbach and Christian Köhler. Health-related searches on the Internet. *JAMA*, 2004.

Emmanuel Farhi and Iván Werning. Insurance and taxation over the life cycle. *Review of Economic Studies*, 80(2):596–635, 2013.

Roberto M Fernandez and Isabel Fernandez-Mateo. Networks, race, and hiring. *American sociological review*, 71(1):42–71, 2006.

Alexander S Fiksdal, Ashok Kumbamu, Ashutosh S Jadhav, Cristian Cocos, Laurie A Nelsen, Jyotishman Pathak, and Jennifer B McCormick. Evaluating the process of online health information searching: a qualitative approach to exploring consumer perspectives. *Journal of Medical Internet research*, 2014.

Aris Filos-Ratsikas, Søren Kristoffer Stiil Frederiksen, and Jie Zhang. Social welfare in one-sided matchings: Random priority and beyond. In *International Symposium on Algorithmic Game Theory*, pages 1–12. Springer, 2014.

Sebastian Findeisen and Dominik Sachs. Optimal need-based financial aid.

Technical report, Mimeo, University of Mannheim/European University Institute, 2016.

Michael H Fox, Ellen Averett, Gail Hansen, and John S Neuberger. The effect of health communications on a statewide west nile virus public health education campaign. *American Journal of Health Behavior*, 30(5):483–494, 2006.

Susannah Fox and Maeve Duggan. Health online 2013. *Washington, DC: Pew Internet & American Life Project*, 1, 2013.

Raoul Fransen-dos Santos. Young people, sexual and reproductive health and HIV. *Bulletin of the WHO*, 87, 2009.

Batya Friedman and David G Hendry. *Value sensitive design: Shaping technology with moral imagination*. Mit Press, 2019.

Batya Friedman and Helen Nissenbaum. Bias in computer systems. *ACM Transactions on Information Systems (TOIS)*, 14(3):330–347, 1996.

Roland G Fryer Jr and Glenn C Loury. Valuing diversity. *Journal of Political Economy*, 121(4):747–774, 2013.

Jugal Garg, Martin Hoefer, and Kurt Mehlhorn. Approximating the Nash social welfare with budget-additive valuations. In *ACM-SIAM Symposium on Discrete Algorithms, SODA '18*, pages 2326–2340, 2018.

Timnit Gebru. Oxford handbook on AI ethics book chapter on race and gender, 2019.

Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumeé III, and Kate Crawford. Datasheets for datasets. *arXiv preprint arXiv:1803.09010*, 2018.

Betsy Gelb and Madeline Johnson. Word-of-mouth communication: Causes and consequences. *Journal of health care marketing*, 15:54–54, 1995.

Becky L Genberg, Zdenek Hlavka, Kelika A Konda, Suzanne Maman, Suwat Chariyalertsak, Alfred Chingono, Jessie Mbwambo, Precious Modiba, Heidi Van Rooyen, and David D Celentano. A comparison of HIV/AIDS-related stigma in four countries: Negative attitudes and perceived acts of discrimination towards people living with HIV/AIDS. *Social science & medicine*, 2009.

Adrienne Germain. Integrating gender into HIV/AIDS programmes in the health sector: tool to improve responsiveness to women's needs. *Bulletin of the WHO*, 2009.

Amira Ghenai and Yelena Mejova. Fake cures: User-centric modeling of health misinformation in social media. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):58, 2018.

Lena Giesbert and Kati Schindler. Assets, shocks, and poverty traps in rural mozambique. *World Development*, 40(8):1594–1609, 2012.

Tarleton Gillespie. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press, 2018.

Tarleton Gillespie and Nick Seaver. Critical algorithm studies: a reading list. `https://socialmediacollective.org/reading-lists/critical-algorithm-studies/`. Accessed: 2019-08-01.

Michele Estrin Gilman. The class differential in privacy law. *Brooklyn Law Review*, 77:1389, 2011.

Jeremy Ginsberg, Matthew H Mohebbi, Rajan S Patel, Lynnette Brammer, Mark S Smolinski, and Larry Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 2009.

Global Fund. Strategic investments for adolescents in HIV, tuberculosis and malaria programs, 2016.

Jacob K Goeree, Arno Riedl, and Aljaž Ule. In search of stars: Network formation among heterogeneous agents. *Games and Economic Behavior*, 67(2), 2009.

Andrew V Goldberg, Jason D Hartline, Anna R Karlin, Michael Saks, and Andrew Wright. Competitive auctions. *Games and Economic Behavior*, 55(2):242–269, 2006.

Mikhail Golosov, Aleh Tsyvinski, Ivan Werning, Peter Diamond, and Kenneth L Judd. New dynamic public finance: A user's guide [with comments and discussion]. *NBER macroeconomics annual*, 21:317–387, 2006.

Belinda Chimphamba Gombachika, Ellen Chirwa, Alfred Maluwa, et al. Sources of information on HIV and sexual and reproductive health for couples living with HIV in rural southern malawi. *AIDS Research and Treatment*, 2013, 2013.

Carla Gomes, Thomas Dietterich, Christopher Barrett, Jon Conrad, Bistra Dilkina, Stefano Ermon, Fei Fang, Andrew Farnsworth, Alan Fern, Xiaoli Fern, Daniel Fink, Douglas Fisher, Alexander Flecker, Daniel Freund, Angela Fuller, John Gregoire, John Hopcroft, Steve Kelling, Zico Kolter, Warren Powell, Nicole Sintov, John Selker, Bart Selman, Daniel Sheldon, David Shmoys, Milind Tambe, Weng-Keen Wong, Christopher Wood, Xiaojian Wu, Yexiang Xue, Amulya Yadav, Abdul-Aziz Yakubu, and Mary Lou Zeeman. Computational sustainability: Computing for a better world and a sustainable future.

Communications of the ACM, 62(9):56–65, August 2019. ISSN 0001-0782. doi: 10.1145/3339399. URL `http://doi.acm.org/10.1145/3339399`.

Mark S Granovetter. The strength of weak ties. In *Social networks*, pages 347–367. Elsevier, 1977.

Ben Green. "fair" risk assessments: A precarious approach for criminal justice reform. In *5th Workshop on Fairness, Accountability, and Transparency in Machine Learning*, 2018.

Daniel Greene, Anna Lauren Hoffmann, and Luke Stark. Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.

David B Grusky. *Social stratification: Class, race, and gender in sociological perspective*. Routledge, 2018.

David B Grusky and Manwai C Ku. Gloom, doom, and inequality. *Social stratification: Class, race, and gender in sociological perspective*, 3:2–28, 2008.

David B. Grusky and Kim A. Weeden. Do disciplinary boundaries keep us from asking the right questions about inequality?, August 2016.

Mingyu Guo and Vincent Conitzer. Worst-case optimal redistribution of VCG payments. In *ACM Conference on Economics and Computation, EC '07*, pages 30–39. ACM, 2007.

Ido Guy. Social recommender systems. In *Recommender Systems Handbook*, pages 511–543. Springer, 2015.

Gregory D Hager, Ann Drobnis, Fei Fang, Rayid Ghani, Amy Greenwald, Terah Lyons, David C Parkes, Jason Schultz, Suchi Saria, Stephen F Smith, et al. Artificial intelligence for social good. *arXiv preprint arXiv:1901.05406*, 2019.

Kevin D Haggerty. Methodology as a knife fight: The process, politics and paradox of evaluating surveillance. *Critical Criminology*, 17(4):277, 2009.

Darrick Hamilton and William Darity Jr. Can 'baby bonds' eliminate the racial wealth gap in putative post-racial america? *The Review of Black Political Economy*, 37(3-4):207–216, 2010.

Denise Hamilton. The truth about DARE; the big-bucks anti-drug program for kids doesn't work. *Los Angeles Times*, 1997.

Anikó Hannák, Claudia Wagner, David Garcia, Alan Mislove, Markus Strohmaier, and Christo Wilson. Bias in online freelance marketplaces: Evidence from taskrabbit and fiverr., 2017.

Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. Strategic classification. In *Proceedings of the 2016 ACM conference on innovations in theoretical computer science*, pages 111–122. ACM, 2016.

Rachel Harmon. Why do we (still) lack data on policing. *Marquette Law Review*, 96:1119, 2012.

Jason Hartline, Vahab Mirrokni, and Mukund Sundararajan. Optimal marketing strategies over social networks. 2008.

Jason D Hartline. Mechanism design and approximation. *Book draft*, 122, October 2013.

Jason D. Hartline and Anna R. Karlin. Profit maximization in mechanism design. In Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V Vazirani, editors, *Algorithmic game theory*, chapter 13. Cambridge University Press Cambridge, 2007.

Jason D. Hartline and Tim Roughgarden. Optimal mechanism design and money burning. In *ACM Symposium on Theory of Computing*, STOC '08, pages 75–84, 2008.

Drew Harwell and Nick Miroff. Ice just abandoned its dream of 'extreme vetting' software that could predict whether a foreign visitor would become a terrorist. *Washington Post*, 2018.

Layla Hasan and Emad Abuelrub. Assessing the quality of web sites. *Applied Computing and Informatics*, 2011.

Nabil Hassein. Against black inclusion in facial recognition, 2017. URL https://digitaltalkingdrum.com/2017/08/15/against-black-inclusion-in-facial-recognition/.

Simon I Hay, Dylan B George, Catherine L Moyes, and John S Brownstein. Big data opportunities for global infectious disease surveillance. *PLoS medicine*, page e1001413, 2013.

Adam Douglas Henry, Paweł Prałat, and Cun-Quan Zhang. Emergence of segregation in evolving social networks. *Proceedings of the National Academy of Sciences*, 108(21), 2011.

Shawndra Hill, Jun Mao, Lyle Ungar, Sean Hennessy, Charles E Leonard, and John Holmes. Natural supplements for H1N1 influenza: retrospective observational infodemiology study of information and search activity on the Internet. *JMIR*, 2011.

Christian Hipp and Michael Plum. Optimal investment for insurers. *Insurance: Mathematics and Economics*, 27(2):215–228, 2000.

Anna Lauren Hoffmann. Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication & Society*, 22(7): 900–915, 2019.

Timothy P Hogan and Carole L Palmer. Information preferences and practices among people living with HIV/AIDS: results from a nationwide survey. *Journal of the Medical Library Association*, 2005.

John J Horton. Price floors and employer preferences: Evidence from a minimum wage experiment. 2017.

John J Horton. Buyer uncertainty about seller capacity: Causes, consequences, and a partial solution. 2018.

John J Horton and Ramesh Johari. At what quality and what price?: Eliciting buyer preferences as a market design problem. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pages 507–507. ACM, 2015.

Keith Hoskin. The 'awful idea of accountability': inscribing people into the measurement of objects. *Accountability: Power, ethos and the technologies of managing*, 265, 1996.

Hadi Hosseini, Kate Larson, and Robin Cohen. Investigating the characteristics of one-sided matching mechanisms under various preferences and risk attitudes. *Autonomous Agents and Multi-Agent Systems*, 32(4):534–567, 2018.

Lily Hu and Yiling Chen. Fairness at equilibrium in the labor market. *arXiv preprint arXiv:1707.01590*, 2017.

Wenjie Hu, Jay Harshadbhai Patel, Zoe-Alanah Robert, Paul Novosad, Samuel Asher, Zhongyi Tang, Marshall Burke, David Lobell, and Stefano Ermon. Mapping missing population in rural India: A deep learning approach with satellite imagery. *arXiv preprint arXiv:1905.02196*, 2019.

Yansong Hu and Christophe Van den Bulte. Nonmonotonic status effects in new product adoption. *Marketing Science*, 33(4):509–533, 2014.

Jevan Hutson, Jessie G Taft, Solon Barocas, and Karen Levy. Debiasing desire: Addressing bias & discrimination on intimate platforms. *arXiv preprint arXiv:1809.01563*, 2018.

Aanund Hylland and Richard Zeckhauser. The efficient allocation of individuals to positions. *Journal of Political economy*, 87(2):293–314, 1979.

IHME. GBD compare data visualization, 2016.

Nicole Immorlica, Brendan Lucier, Glen Weyl, and Joshua Mollner. Approximate efficiency in matching markets. In *International Conference on Web and Internet Economics*, pages 252–265. Springer, 2017.

Lucas D Introna and Helen Nissenbaum. Shaping the web: Why the politics of search engines matters. *The information society*, 16(3):169–185, 2000.

Sanae Ito. Microfinance and social capital: does social capital help create good practice? *Development in Practice*, 13(4):322–332, 2003.

ITU. International Telecommunication Union statistics, 2017.

Matthew O. Jackson and Brian W. Rogers. Meeting strangers and friends of friends: How random are social networks? *American Economic Review*, 97(3): 890–915, June 2007. doi: 10.1257/aer.97.3.890.

Matthew O. Jackson, Tomas Rodriguez-Barraquer, and Xu Tan. Social capital and social quilts: Network patterns of favor exchange. *American Economic Review*, 102(5), May 2012. doi: 10.1257/aer.102.5.1857.

Gregor Jarosch. Searching for job security and the consequences of job loss. *manuscript, Stanford University*, 2015.

Sheila Jasanoff. Technology as a site and object of politics. In *The Oxford handbook of contextual political analysis*. Oxford University Press, 2006.

Neal Jean, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon. Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301), 2016. ISSN 0036-8075. doi: 10.1126/science.aaf7894. URL `http://science.sciencemag.org/content/353/6301/790`.

Will Jones and Alexander Teytelboym. Matching systems for refugees. *Journal on Migration and Human Security*, 5:667–681, 2018.

Yogesh V. Joshi, David J. Reibstein, and Z. John Zhang. Turf wars: Product line strategies in competitive markets. *Marketing Science*, 35(1):128–141, 2016.

Steve Jurvetson. What exactly is viral marketing? 78, 2000.

Michael L. Katz and Carl Shapiro. Network externalities, competition, and compatibility. 75(3):424–440, June 1985.

David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, pages 137–146, New York, NY, USA, 2003. ACM. ISBN 1-58113-737-0. doi: 10.1145/956750.956769. URL `http://doi.acm.org/10.1145/956750.956769`.

Margaret L Kern, Gregory Park, Johannes C Eichstaedt, H Andrew Schwartz, Maarten Sap, Laura K Smith, and Lyle H Ungar. Gaining insights from social media language: Methodologies and challenges. *Psychological methods*, 21(4): 507, 2016.

Yoko Kijima, Tomoya Matsumoto, and Takashi Yamano. Nonfarm employment, agricultural shocks, and poverty dynamics: evidence from rural Uganda. *Agricultural Economics*, 35:459–467, 2006.

Kibae Kim and Jörn Altmann. Effect of homophily on network formation. *Communications in Nonlinear Science and Numerical Simulation*, 44, 2017.

Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. Inherent tradeoffs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807*, 2016.

Bart P Knijnenburg, Saadhika Sivakumar, and Daricia Wilkinson. Recommender systems for self-actualization. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 11–14. ACM, 2016.

Anjini Kochar. Explaining household vulnerability to idiosyncratic income shocks. *The American Economic Review*, 85(2):159–164, 1995.

Anjini Kochar. Smoothing consumption by smoothing income: hours-of-work responses to idiosyncratic agricultural shocks in rural india. *Review of Economics and Statistics*, 81(1):50–61, 1999.

Narayana R Kocherlakota. Wedges and taxes. *American Economic Review*, 94(2): 109–113, 2004.

John Logan Koepke and David G Robinson. Danger ahead: Risk assessment and the future of bail reform. *Washington Law Review*, 93:1725, 2018.

Fuhito Kojima and Mihai Manea. Incentives in the probabilistic serial mechanism. *Journal of Economic Theory*, 145(1):106 – 123, 2010. ISSN 0022-0531.

Scott Duke Kominers. Good markets (really do) make good neighbors. *ACM SIGecom Exchanges*, 16(2):12–26, 2018.

Gueorgi Kossinets and Duncan J Watts. Empirical analysis of an evolving social network. *science*, 311(5757):88–90, 2006.

Gueorgi Kossinets and Duncan J Watts. Origins of homophily in an evolving social network. *American journal of sociology*, 115(2):405–450, 2009.

Balzs Kovcs and Amanda J. Sharkey. The paradox of publicity: How awards can negatively affect the evaluation of quality. *Administrative Science Quarterly*, 59 (1):1–33, 2014.

Steven Krauwer. The basic language resource kit (blark) as the first milestone for the language resources roadmap. *Proceedings of SPECOM 2003*, pages 8–15, 2003.

Amanda R Kube, Sanmay Das, and Patrick J. Fowler. Allocating interventions based on counterfactual predictions: A case study on homelessness services. 2018.

Alexandra V Kulinkina, Karen C Kosinski, Alexander Liss, Michael N Adjei, Gilbert A Ayamgah, Patrick Webb, David M Gute, Jeanine D Plummer, and Elena N Naumova. Piped water consumption in ghana: A case study of temporal and spatial patterns of clean water demand relative to alternative water sources in rural small towns. *Science of the Total Environment*, 559:291–301, 2016.

Smita Kumar and Kristin Mmari. Programming considerations for youth-friendly HIV care and treatment services. *From the Ground Up*, 2004.

Matt Laslo. The CDC could totally study gun violence-it just needs money. *Wired*, 2019.

Paul F Lazarsfeld, Robert K Merton, et al. Friendship as a social process: A substantive and methodological analysis. *Freedom and control in modern society*, 18(1):18–66, 1954.

Leadership Conference on Civil and Human Rights. The use of pretrial "risk assessment" instruments: A shared statement of civil rights concerns, 2018. URL `http://civilrightsdocs.info/pdf/criminal-justice/Pretrial-Risk-Assessment-Full.pdf`.

Vicki Lens, Margo Nugent, Christopher Wimer, and Antonina Pavlenko. Poverty tracker – monitoring poverty and well-being in NYC: Report 5. Technical report, Columbia Population Research Center and Robin Hood, 2016.

Jacob Leshno. Dynamic matching in overloaded waiting lists. 2017.

Jure Leskovec, Lada Adamic, and Bernardo Huberman. The dynamics of viral marketing. 1(1), May 2007a.

Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1), March 2007b. ISSN 1556-4681. doi: 10.1145/1217299.1217301. URL `http://doi.acm.org/10.1145/1217299.1217301`.

Lawrence Lessig. *Code: And other laws of cyberspace*. Basic Books, 2009.

Karen Levy and Solon Barocas. Designing against discrimination in online markets. *Berkeley Technology Law Journal*, 32:1183, 2017.

Xiaoming Li, Chongde Lin, Zuxin Gao, Bonita Stanton, Xiaoyi Fang, Qin Yin, and Ying Wu. HIV/AIDS knowledge and the implications for health promotion programs among chinese college students: geographic, gender and age differences. *HPI*, 2004.

Nan Lin. Social networks and status attainment. *Annual review of sociology*, 25 (1):467–487, 1999.

Rebecca Ling and Joon Lee. Disease monitoring and health campaign evaluation using google search activities for HIV and AIDS, stroke, colorectal cancer, and marijuana use in canada: a retrospective observational study. *JMIR*, 2016.

Leslie S Liu, Jina Huh, Tina Neogi, Kori Inkpen, and Wanda Pratt. Health vlogger-viewer interaction in chronic illness management, 2013.

Irene Lo. Mechanism design and allocation systems in education. 2017.

Emma Lurie and Eni Mustafaraj. Opening up the black box: Auditing google's top stories algorithm. In *Proceedings of the International Florida Artificial Intelligence Research Society Conference*, volume 32, 2019.

Robert Manduca. Mechanism design for social good. *RobertManduca.com*, 2019. URL `http://robertmanduca.com/portfolio/mechanism-design-4-social-good/`.

Christopher Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to information retrieval*, volume 16. Cambridge university press, 2010.

Cynthia M Marlowe, Sandra L Schneider, and Carnot E Nelson. Gender and attractiveness biases in hiring decisions: Are more experienced managers less biased? *Journal of applied psychology*, 81(1):11, 1996.

Alice Marwick, Claire Fontaine, and danah boyd. "nobody sees it, nobody gets mad": Social media, privacy, and personal responsibility among low-ses youth. *Social Media+ Society*, 3(2):2056305117710455, 2017.

Adalbert Mayer and Steven L Puller. The old boy (and girl) network: Social network formation on university campuses. *Journal of Public Economics*, 92 (1-2), 2008.

Sandra G. Mayson. Bias in, bias out. *Yale Law Journal*, 128:2218, 2019.

Ngozi C Mbonu, Bart van den Borne, and Nanne K De Vries. Stigma of people with HIV/AIDS in sub-Saharan Africa: a literature review. *Journal of Tropical Medicine*, 2009.

Andrew Kachites McCallum. MALLET: A machine learning for language toolkit. http://mallet.cs.umass.edu, 2002.

Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1), 2001.

Steven Mello. Speed trap or poverty trap? fines, fees, and financial wellbeing. 2018.

James A Mirrlees. An exploration in the theory of optimum income taxation. *The review of economic studies*, 38(2):175–208, 1971.

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 220–229. ACM, 2019.

Thandika Mkandawire. *Targeting and universalism in poverty reduction*. United Nations Research Institute for Social Development Geneva, 2005.

Jonathan Morduch. Poverty and vulnerability. *The American Economic Review*, 84(2):221–225, 1994.

Hervé Moulin. *Fair Division and Collective Welfare*. 2003.

Hervé Moulin. Almost budget-balanced VCG mechanisms to assign multiple objects. *Journal of Economic theory*, 144(1):96–119, 2009.

Lev Muchnik, Sinan Aral, and Sean J Taylor. Social influence bias: A randomized experiment. *Science*, 341(6146):647–651, 2013.

Daniel Mutembesa, Christopher Omongo, and Ernest Mwebaze. Crowdsourcing real-time viral disease and pest information: A case of nation-wide cassava disease surveillance in a developing country. In *HCOMP*, pages 117–125, 2018.

Ernest Mwebaze and Michael Biehl. Prototype-based classification for image analysis and its application to crop disease diagnosis. In *Advances in Self-*

Organizing Maps and Learning Vector Quantization, pages 329–339. Springer, 2016.

Ernest Mwebaze and Godliver Owomugisha. Machine learning for plant disease incidence and severity measurements from leaf images. In *Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on*, pages 158–163. IEEE, 2016.

J. Nash. The bargaining problem. *Econometrica*, 18(2):155–162, April 1950.

Mark EJ Newman. Assortative mixing in networks. *Physical review letters*, 89 (20):208701, 2002.

Neil Newman, Lauren Falcao Bergquist, Nicole Immorlica, Kevin Leyton-Brown, Brendan Lucier, Craig McIntosh, John Quinn, and Richard Ssekibuule. Designing and evolving an electronic agricultural marketplace in Uganda. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, page 14. ACM, 2018.

Albert L Nichols and Richard J Zeckhauser. Targeting transfers through restrictions on recipients. *The American Economic Review*, 72(2):372–377, 1982.

Shirin Nilizadeh, Anne Groggel, Peter Lista, Srijita Das, Yong-Yeol Ahn, Apu Kapadia, and Fabio Rojas. Twitter's glass ceiling: The effect of perceived gender on online visibility., 2016.

Noam Nisan and Amir Ronen. Algorithmic mechanism design. *Games and Economic behavior*, 35(1-2):166–196, 2001.

William Nnyanja. The role of saving and credit cooperatives in improving household income: The case for teachers in mukono district, 2017.

Safiya Umoja Noble. *Algorithms of oppression: How search engines reinforce racism*. NYU Press, 2018.

Michelle Obama. *Becoming*. Crown Publishing, 2018.

Ziad Obermeyer and Sendhil Mullainathan. Dissecting racial bias in an algorithm that guides health decisions for 70 million people. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 89–89. ACM, 2019.

Alex J Ocampo, Rumi Chunara, and John S Brownstein. Using search queries for malaria surveillance, Thailand. *Malaria journal*, 2013.

Brendan O'Flaherty. What shocks precipitate homelessness. *Unpublished manuscript. New York: Columbia University*, 2009.

Laura O'Grady. Meeting health information needs of people with HIV/AIDS: sources and means of collaboration. *Health Information & Libraries Journal*, 2008.

Cathy O'neil. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books, 2016.

Abraham Othman, Tuomas Sandholm, and Eric Budish. Finding approximate competitive equilibria: Efficient and fair course allocation. In *International Conference on Autonomous Agents and Multiagent Systems*, pages 873–880, 2010.

Abraham Othman, Christos Papadimitriou, and Aviad Rubinstein. The complexity of fairness through equilibrium. *ACM Transactions on Economics and Computation (TEAC)*, 4(4):20, 2016.

Rebekah Overdorf, Bogdan Kulynych, Ero Balsa, Carmela Troncoso, and Seda Gürses. Questioning the assumptions behind fairness solutions. *arXiv preprint arXiv:1811.11293*, 2018.

Scott E. Page. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton University Press, 2008.

Bo Pang and Lillian Lee. *Opinion Mining and Sentinment Analysis*. Number 2(1-2) in Foundations and Trends in Information Retrieval. 2008.

Alula Pankhurst and Damen Haile Mariam. The "iddir" in ethiopia: Historical development, social function, and potential role in hiv/aids prevention and control. *Northeast African Studies*, pages 35–57, 2000.

Pietro Panzarasa, Tore Opsahl, and Kathleen M. Carley. Patterns and dynamics of users' behavior and interaction: Network analysis of an online community. *Journal of the American Society for Information Science and Technology*, 60(5):911–932, 2009.

Ashwin Paranjape, Austin R. Benson, and Jure Leskovec. Motifs in temporal networks. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, WSDM '17, pages 601–610, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4675-7. doi: 10.1145/3018661.3018731. URL `http://doi.acm.org/10.1145/3018661.3018731`.

Frank Pasquale. *The black box society*. Harvard University Press, 2015.

Frank Pasquale. Odd numbers. *Real Life Mag*, 2018.

Samir Passi and Solon Barocas. Problem formulation and fairness. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 39–48. ACM, 2019.

Neil Patel, Deepti Chittamuru, Anupam Jain, Paresh Dave, and Tapan S Parikh. Avaaj otalo: a field study of an interactive voice forum for small farmers in rural india. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 733–742. ACM, 2010.

Parag A. Pathak and Jay Sethuraman. Lotteries in student assignment: An equivalence result. *Theoretical Economics*, 6(1):1–17, 2011. ISSN 1555-7561.

Michael J. Paul and Mark Dredze. Social monitoring for public health. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 2017.

David S Pedulla and Devah Pager. Race and networks in the job search process. *American Sociological Review*, page 0003122419883255, 2017.

Seeta Peña Gangadharan and Jedrzej Niklas. Decentering technology in discourse on discrimination. *Information, Communication & Society*, 22(7):882–899, 2019.

Scott R Peppet. Unraveling privacy: The personal prospectus and the threat of a full-disclosure future. *Northwestern University Law Review*, 105:1153, 2011.

Govind Persad, Alan Wertheimer, and Ezekiel J Emanuel. Principles for allocation of scarce medical interventions. *The Lancet*, 373(9661):423–431, 2009.

Thomas Piketty. *The economics of inequality*. Harvard University Press, 2015.

Philip M Polgreen, Yiling Chen, David M Pennock, Forrest D Nelson, and Robert A Weinstein. Using Internet searches for influenza surveillance. *Clinical infectious diseases*, 2008.

Julia Powles. The seductive diversion of 'solving' bias in artificial intelligence, 2018.

Ariel D. Procaccia. Cake cutting: Not just child's play. *Communications of the ACM*, 56(7):78–87, July 2013. ISSN 0001-0782.

Ariel D. Procaccia and Hervé Moulin. *Cake Cutting Algorithms*, page 311330. Cambridge University Press, 2016. doi: 10.1017/CBO9781107446984.014.

Ruchir Puri. Mitigating bias in ai models. *IBM Research Blog*, 2018.

John Alexander Quinn, Kevin Leyton-Brown, and Ernest Mwebaze. Modeling and monitoring crop disease in developing countries. In *AAAI*, 2011.

Raj Jayadev. The future of pretrial justice is not bail or system supervision — it is freedom and community. *Silcon Valley Debug*, 2019.

Inioluwa Deborah Raji and Joy Buolamwini. Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. *AAAI/ACM Conference on AI Ethics and Society*, 2019.

Anatol Rapoport. Spread of information through a population with socio-structural bias: I. assumption of transitivity. *The bulletin of mathematical biophysics*, 15(4):523–533, 1953.

Joel R Reidenberg. Lex informatica: The formulation of information policy rules through technology. *Texas Law Review*, 76:553, 1997.

Paul Resnick, Richard Zeckhauser, John Swanson, and Kate Lockwood. The value of reputation on eBay: A controlled experiment. *Experimental Economics*, 9:79–101, 2006.

Matthew Richardson and Pedro Domingos. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '02, pages 61–70, New York, NY, USA, 2002. ACM. ISBN 1-58113-567-X. doi: 10.1145/775047.775057. URL `http://doi.acm.org/10.1145/775047.775057`.

John Roach. Microsoft improves facial recognition technology to perform well across all skin tones, genders. *The AI Blog*, 2018.

Sarah T Roberts. *Behind the Screen: Content Moderation in the Shadows of Social Media*. Yale University Press, 2019.

Jack Robertson and William Webb. *Cake Cutting Algorithms: Be Fair If You Can*. 1998.

Everett Rogers. *Diffusion of Innovations*. Free Press, fourth edition, 1995.

David Rolnick, Priya L Donti, Lynn H Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, et al. Tackling climate change with machine learning. *arXiv preprint arXiv:1906.05433*, 2019.

Robin Root. Situating experiences of hiv-related stigma in swaziland. *Global public health*, 2010.

Rahmtin Rotabi, Krishna Kamath, Jon Kleinberg, and Aneesh Sharma. Cascades: A view from the audience. 2017.

Alvin E Roth. Repugnance as a constraint on markets. *Journal of Economic perspectives*, 21(3):37–58, 2007.

Alvin E Roth. What have we learned from market design? *Innovations: Technology, Governance, Globalization*, 3(1):119–147, 2008.

Alvin E. Roth, Tayfun Sönmez, and M. Utku Unver. Kidney exchange. *The Quarterly Journal of Economics*, 119(2), 2004. doi: 10.1162/0033553041382157. URL +http://dx.doi.org/10.1162/0033553041382157.

Alvin E Roth, Tayfun Sönmez, et al. A kidney exchange clearinghouse in new england. *American Economic Review*, 95(2):376–380, 2005.

Kenneth J Rothman, Sander Greenland, Timothy L Lash, et al. Modern epidemiology. 3, 2008.

Emmanuel Saez and Stefanie Stantcheva. A simpler theory of optimal capital taxation. *Journal of Public Economics*, 162:120–142, 2018.

Nithya Sambasivan, Ed Cutrell, Kentaro Toyama, and Bonnie Nardi. Intermediated technology use in developing communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 2583–2592, New York, NY, USA, 2010. ACM.

Luis Gomes Sambo. *The health of the people: what works: the African Regional Health Report 2014*. World Health Organization, 2014.

Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. Auditing algorithms: Research methods for detecting discrimination on Internet platforms. *Data and discrimination: converting critical concerns into productive inquiry*, 22, 2014.

Mauricio Santillana, André T Nguyen, Mark Dredze, Michael J Paul, Elaine O Nsoesie, and John S Brownstein. Combining search, social media, and traditional data sources to improve influenza surveillance. *PLoS computational biology*, 2015.

Rofiah O Sarumi and Ann E Strode. New law on HIV testing in botswana: The implications for healthcare professionals. *Southern African Journal of HIV Medicine*, 2015.

Hanspeter Schmidli et al. On minimizing the ruin probability by investment and reinsurance. *The Annals of Applied Probability*, 12(3):890–907, 2002.

Tobias Schnabel, Paul N Bennett, Susan T Dumais, and Thorsten Joachims. Short-term satisfaction and long-term coverage: Understanding how users tolerate algorithmic exploration. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 513–521. ACM, 2018.

H. Andrew Schwartz and Lyle H Ungar. Data-driven content analysis of social media: A systematic overview of automated methods. *The Annals of the American Academy of Political and Social Science*, 659, 2015.

H. Andrew Schwartz, Johannes C. Eichstaedt, Margaret L. Kern, Lukasz Dziurzynski, Stephanie M. Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin E. P. Seligman, and Lyle H. Ungar. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLOS ONE*, 2013.

H. Andrew Schwartz, Salvatore Giorgi, Maarten Sap, Patrick Crutchley, Lyle Ungar, and Johannes Eichstaedt. DLATK: Differential language analysis toolkit, 2017.

Alon Sela, Erez Shmueli, Dima Goldenberg, and Irad Ben-Gal. Why spending more might get you less, dynamic selection of influencers in social networks. In *2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE)*, pages 1–4, Nov 2016. doi: 10.1109/ICSEE.2016.7806079.

Andrew D Selbst, Danah Boyd, Sorelle A Friedler, Suresh Venkatasubramanian, and Janet Vertesi. Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 59–68. ACM, 2019.

Amartya Sen. Poverty: an ordinal approach to measurement. *Econometrica: Journal of the Econometric Society*, pages 219–231, 1976.

Amartya Sen, James E Foster, et al. *On economic inequality*. Oxford University Press, 1997.

Thomas M Shapiro et al. *The hidden cost of being African American: How wealth perpetuates inequality*. Oxford University Press, USA, 2004.

Kristen S Shuyler and Kristin M Knight. What are patients seeking when they turn to the Internet? qualitative content analysis of questions asked by visitors to an orthopaedics web site. *Journal of Medical Internet research*, 2003.

Elizabeth Sillence, Pam Briggs, Peter Richard Harris, and Lesley Fishwick. How do patients evaluate and make use of online health information? *Social science & medicine*, 2007.

Mwiza Simbeye Simbeye, Mwila Kangwa, and Patrick Sikalinda. Bridging markets for small scale farmers with mechanism design, 2019.

Theda Skocpol. Targeting within universalism: Politically viable policies to combat poverty in the united states. *The urban underclass*, 411(411):437–59, 1991.

Amanda Spink, Yin Yang, Jim Jansen, Pirrko Nykanen, Daniel P Lorence, Seda Ozmutlu, and H Cenk Ozmutlu. A study of medical and health queries to web search engines. *Health Information & Libraries Journal*, 2004.

Jeff Spross. How robots became a scapegoat for the destruction of the working class. *The Week*, 2019. URL `https://theweek.com/articles/837759/how-robots-became-scapegoat-destruction-working-class`.

Richard Ssekibuule, John A Quinn, and Kevin Leyton-Brown. A mobile market for agricultural trade in Uganda. In *Proceedings of the 4th Annual Symposium on Computing for Development*, page 9. ACM, 2013.

Christoph Stadtfeld. The micro–macro link in social networks. *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*, pages 1–15, 2015.

Ana-Andreea Stoica, Christopher Riederer, and Augustin Chaintreau. Algorithmic glass ceiling in social networks: The effects of social recommendations on network diversity, 2018.

Samantha Stonbraker, Montina Befus, Leonel Lerebours Nadal, Mina Halpern, and Elaine Larson. Factors associated with health information seeking, processing, and use among HIV positive adults in the dominican republic. *AIDS and Behavior*, 2017.

Jessica Su, Aneesh Sharma, and Sharad Goel. The effect of recommendations on network structure, 2016.

Sharifa Sultana, François Guimbretière, Phoebe Sengers, and Nicola Dell. Design within a patriarchal society: Opportunities and challenges in designing for rural women in bangladesh. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 536:1–536:13. ACM, 2018.

Latanya Sweeney. Discrimination in online ad delivery. *ACM Queue*, 2013.

Neil Thakral. The public-housing allocation problem. Technical report, Technical report, Harvard University, 2016.

William Thomson. The Fair Division of a Fixed Supply Among a Growing Population. *Mathematics of Operations Research*, 8(3):319–326, August 1983.

William Thomson and Terje Lensberg. *Axiomatic Theory of Bargaining with a Variable Number of Agents*. Cambridge University Press, 1989. doi: 10.1017/CBO9780511664489.

Gergő Tóth, Johannes Wachs, Riccardo Di Clemente, Ákos Jakobi, Bence Ságvári, János Kertész, and Balázs Lengyel. Inequality is rising where social network segregation interacts with urban topology. *arXiv preprint arXiv:1909.11414*, 2019.

Melvin M. Tumin, David Grusky, and Manwai Candy Ku. Gloom, doom, and inequality. 2012.

UNDP. UNDP discussion paper: Gender and malaria, 2015a.

UNDP. UNDP discussion paper: Gender and TB, 2015b.

Hal R Varian. Redistributive taxation as social insurance. *Journal of public Economics*, 14(1):49–68, 1980.

Aditya Vashistha, Edward Cutrell, Gaetano Borriello, and William Thies. Sangeet swara: A community-moderated voice forum in rural india. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 417–426. ACM, 2015a.

Aditya Vashistha, Edward Cutrell, Nicola Dell, and Richard Anderson. Social media platforms for low-income blind people in india. In *Proceedings of the*

17th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '15, pages 259–272. ACM, 2015b.

Aditya Vashistha, Fabian Okeke, Richard Anderson, and Nicola Dell. 'you can always do better!: The impact of social proof on participant response bias. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 552. ACM, 2018.

Vijay V. Vazirani. Combinatorial algorithms for market equilibrium. In Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V Vazirani, editors, *Algorithmic game theory*, chapter 5, pages 103–134. Cambridge University Press Cambridge, 2007.

Vijay V Vazirani. *Approximation algorithms*. Springer Science & Business Media, 2013.

Melanie Wakefield, Brian Flay, Mark Nichter, and Gary Giovino. Effects of anti-smoking advertising on youth smoking: A review. *Journal of Health Communication*, 8:229–247, 2003.

Michael S Wald and Maria Woolverton. Risk assessment: The emperor's new clothes? *Child Welfare: Journal of Policy, Practice, and Program*, 1990.

Daniel Waldinger. Targeting in-kind transfers through market design: A revealed preference analysis of public housing allocation. 2017.

Jianming Wang, Yang Fei, Hongbing Shen, and Biao Xu. Gender difference in knowledge of tuberculosis and associated health-care seeking behaviors: a cross-sectional study in a rural area of china. *BMC Public Health*, 2008.

Nathan Weixel. Funding bill won't prompt new CDC gun research, experts say. *The Hill*, 2018.

Meredith Whittaker, Kate Crawford, Roel Dobbe, Genevieve Fried, Elizabeth Kaziunas, Varoon Mathur, Sarah Mysers West, Rashida Richardson, Jason Schultz, and Oscar Schwartz. *AI Now report 2018*. AI Now Institute at New York University, 2018.

WHO. Integrating gender into HIV/AIDS programmes in the health sector: tool to improve responsiveness to women's needs. Technical report, 2009.

Christopher Wimer, Claudette Bannerman, Irwin Garfinkel, and Kathryn Neckerman. Poverty tracker – monitoring poverty and well-being in NYC: Report 2. Technical report, Columbia Population Research Center and Robin Hood, 2014a.

Christopher Wimer, Irwin Garfinkel, Madeleine Gelblum, Narayani Lasala, Stephanie Phillips, Yajuan Si, Julien Teitler, and Jane Waldfogel. Poverty tracker – monitoring poverty and well-being in NYC: Report 1. Technical report, Columbia Population Research Center and Robin Hood, 2014b.

Christopher Wimer, Sophie Collyer, Irwin Garfinkel, Matthew Maury, Kathryn Neckerman, Julien Teitler, Jane Waldfogel, and David Wutchiett. Poverty tracker – monitoring poverty and well-being in NYC: Report 4. Technical report, Columbia Population Research Center and Robin Hoods, 2016a.

Christopher Wimer, Irwin Garfinkel, Matthew Maury, Kathryn Neckerman, Stephanie Phillips, Julien Teitler, Jane Waldfogel, and David Wutchiett. Poverty tracker – monitoring poverty and well-being in NYC: Report 3. Technical report, Columbia Population Research Center and Robin Hood, 2016b.

Langdon Winner. Do artifacts have politics? *Daedalus*, pages 121–136, 1980.

Michael Xie, Neal Jean, Marshall Burke, David Lobell, and Stefano Ermon. Transfer learning from deep features for remote sensing and poverty mapping. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

Yexiang Xue, Ian Davies, Daniel Fink, Christopher Wood, and Carla P. Gomes. Avicaching: A two stage game for bias reduction in citizen science. In *Proceedings of the 2016 International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '16, pages 776–785, 2016a.

Yexiang Xue, Ian Davies, Daniel Fink, Christopher Wood, and Carla P. Gomes. Behavior identification in two-stage games for incentivizing citizen science exploration. In Michel Rueher, editor, *Principles and Practice of Constraint Programming*, pages 701–717, Cham, 2016b. Springer International Publishing.

Amulya Yadav, Hau Chan, Albert Xin Jiang, Haifeng Xu, Eric Rice, and Milind Tambe. Using social networks to aid homeless shelters: Dynamic influence maximization under uncertainty. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, pages 740–748. International Foundation for Autonomous Agents and Multiagent Systems, 2016.

Amulya Yadav, Bryan Wilder, Eric Rice, Robin Petering, Jaih Craddock, Amanda Yoshioka-Maxwell, Mary Hemler, Laura Onasch-Vera, Milind Tambe, and Darlene Woo. Influence maximization in the field: The arduous journey from emerging to deployed application. In *Proceedings of the 16th conference on autonomous agents and multiagent systems*, pages 150–158. International Foundation for Autonomous Agents and Multiagent Systems, 2017.

Amulya Yadav, Bryan Wilder, Eric Rice, Robin Petering, Jaih Craddock, Amanda Yoshioka-Maxwell, Mary Hemler, Laura Onasch-Vera, Milind

Tambe, and Darlene Woo. Bridging the gap between theory and practice in influence maximization: Raising awareness about HIV among homeless youth. In *IJCAI*, pages 5399–5403, 2018.

Shihao Yang, Mauricio Santillana, and Samuel C Kou. Accurate estimation of influenza epidemics using google search data via argo. *Proceedings of the National Academy of Sciences*, 2015.

Lakshman Yapa. What causes poverty?: A postmodern view. *Annals of the Association of American Geographers*, 86(4):707–728, 1996.

H. Peyton Young. *Equity: In Theory and Practice*. 1995.

Qingyu Yuan, Elaine O Nsoesie, Benfu Lv, Geng Peng, Rumi Chunara, and John S Brownstein. Monitoring influenza epidemics in china with search query from baidu. *PloS one*, 2013.

Andrey Zheluk, Casey Quinn, Daniel Hercz, and James A Gillespie. Internet search patterns of human immunodeficiency virus and the digital divide in the russian federation: infoveillance study. *Journal of medical Internet research*, 2013.

Lin Zhou. On a conjecture by Gale about one-sided matching problems. *Journal of Economic Theory*, 52(1):123–135, 1990.

Xi-chuan Zhou and Hai-bin Shen. Notifiable infectious disease surveillance with data collected by search engine. *Journal of Zhejiang University Science C*, 2010.

Philip G Zimbardo and Michael R Leippe. *The psychology of attitude change and social influence.* Mcgraw-Hill Book Company, 1991.