Fairness in Diffusion Processes

Jeffrey Sachs[1]

December 7, 2023

Abstract

This research investigates fairness in information spread within 14 distinct Ugandan villages, focusing on the distribution of information about a political participation technology. The study applies a Simulated Method of Moments approach to assess the impact of different network metrics and fairness definitions on the spread of information across male and female villagers. Initial findings highlight significant differences in the likelihood of achieving fair outcomes with various interventions. Although some fair strategies show slight improvements, they still exhibit low probabilities of fair outcomes. The introduction of a new fairness definition, *Seed Average x Coverage Fairness*, enhances the prediction of fair outcomes across many network metrics, albeit with a trade-off in the efficiency of information spread. The research underscores the difficulty of controlling outcomes in a complex system, and the dynamics of efficiency and fairness, demonstrating that context plays a vital role for these processes.

[1] Department of Sociology, Yale University (corresponding author, jeffrey.sachs@yale.edu) [†]Department of Sociology, Yale University

1 Introduction

Social network research has demonstrated the importance of relationships and relational patterns in directing flows of information and resources. Individuals occupying brokerage positions within networks reap significant benefits (Burt, 1995; 2007). Access to weak ties has positive impacts (Granovetter, 1973). Upward mobility is tied to economic connectedness (Chetty et al., 2022).

Groups as well as individuals can benefit from distinct network topographies. Small world structures with local clustering and random ties have high levels of connectivity which can increase information flow within networks (Watts, 1999). Well-connected community institutions can broker social capital and improve well-being (Small, 2010). Community cohesion supports mutual understanding and can give rise to collective efficacy (Centola, 2015; Small, 2004).

An important implication of this work has been the identification of potential trade-offs between positional benefits for individuals and advantages that accrue to the group. There is long standing tension in social capital research between advantages related to position, which require sparsity, and advantages related to group structure, which require density and cohesion (Burt, 2007; Coleman, 1988; Sorenson et al., 2006; van de Rijt et al., 2009).

The potential benefits of diffusion processes for groups have led researchers to engage in targeted interventions intended to introduce resources or information to communities through diffusion processes (Valente, 2012; Valente and Pitts, 2017). These interventions have been explored across a variety of domains, including: suicide prevention (Pickering et al., 2018), anti-smoking and e-cigarette prevention campaigns (Chu et al., 2021; Cobb et al., 2016; Valente et al., 2003), HIV prevention (Amirkhanian et al., 2005; Kelly et al., 2000; Kelly et al., 1997; Kelly et al., 2006; Yadav et al., 2018), substance abuse prevention (Valente et al., 2007), sexual violence prevention (Edwards et al., 2022), health behavior (Centola, 2011; Centola and van de Rijt, 2015; Christakis and Fowler, 2013; Kim et al., 2015; Zhang et al., 2016; Zhang et al., 2015a; 2015b; Zhang and Centola, 2019), new product/technology adoption (Aral and Walker, 2011; Arslan et al., 2022; Carmel et al., 2009; Chatterjee, 2011; Cheng, 2022; Ferrali et al., 2020; Hanson and Putler, 1996; Iyengar et al., 2015; Iyengar et al., 2011; Peres et al., 2010; Van den Bulte and Joshi, 2007), viral marketing (Hinz et al., 2011; Leskovec et al., 2007; Wang and Street, 2018), public opinion formation (Flaxman et al., 2016; Valente and Davis, 1999; Valente and Pumpuang, 2007; Watts and Dodds, 2007), microfinance (Banerjee et al., 2013), and misinformation (Young et al., 2021), among other settings.

Because of the importance of networks for distributing resources, social network structure is linked to the generation and maintenance of inequality. The process of network diffusion has the potential for increasing inequality by compounding advantages for already advantaged groups and individuals through at least three different mechanisms (DiMaggio and Garip, 2011; 2012). Local network externalities can magnify benefits based on the number of other adopters in a group or ego-network. Social learning can increase the flow of expertise and useful information within already advantaged groups or well-connected individuals. And normative influence operating through relational sanctioning can encourage good behavior and discourage negative behaviors, again benefiting some individuals and groups more than others depending on their network position and the structure of group relations. Other mechanisms may also be operating, such as discrimination which can prevent the formation of beneficial social ties (Phelan and Link, 2015). For social networks in which members of certain groups face barriers to social capital — e.g., settings in which group membership maps onto lower status, like SES, race, or gender — groups experience both differential treatment and disparate impacts in the spread of resources (Pager and Shepherd, 2008).

Targeted network interventions raise issues for the diffusion literature that intersect with the growing literature on fairness in algorithms and AI (Dwork et al., 2012; Hardt et al., 2016). Researchers in science and technology studies and the data sciences field have been identifying ways in which algorithms and procedures embedded in new technologies can exacerbate existing inequalities and injustice in society (Berk et al., 2018; Celis et al., 2019; Lambrecht and Tucker, 2018). Among other important contributions, these researchers have identified a surprising trade-off between ameliorating bias against individuals and ameliorating bias against groups. In particular, optimizing fairness for individuals can reduce fairness for groups and optimizing fairness for groups can reduce fairness for individuals (Corbett-Davies et al., 2017; Kleinberg et al., 2016; Pleiss et al., 2017). And importantly, researchers have shown that small design choices can have big impacts that if unchecked can undermine the intentions of researchers and designers by increasing harmful inequalities rather than mitigate harm or improve life conditions (Benjamin, 2019a; 2019b; Browne, 2015; Eubanks, 2011; 2018; Nelson, 2016).

In this manuscript, we consider how diffusion strategies affect the generation of inequalities between groups and individuals and methods for minimizing bias at the same time. We explore the impact of seeding strategies on information diffusion across 14 independent Ugandan villages, specifically focusing on disparities between male and female villagers. Utilizing a calibrated diffusion model based on real-case scenarios, we assess the effectiveness of a large set of interventions and definitions for fairness. Our initial findings reveal that a randomly chosen fair intervention is not likely to result in fair outcomes, with Betweenness slightly improving the likelihood of achieving equitable outcomes. Despite these observations, the subset of fair interventions that can be chosen, such that our strategy is not randomly selected, exhibits even lower probabilities of fair outcomes. To address this problem, we introduce a fairness definition that significantly improves the potential for outcome fairness across several network metrics, albeit with a noted decrease in the efficiency of information spread, highlighting a potential trade-off between efficiency and fairness in diffusion settings.

The results illuminate both the structural mechanisms through which diffusion processes may affect inequality in general and provide useful information to researchers attempting to make targeted interventions in ways that alleviate rather than exacerbate unequal outcomes.

1.1 Algorithmic Bias and Fairness

In his influential book, A Theory of Justice, John Rawls defined social justice as “a standard whereby the distributive aspects of the basic structure of society are to be assessed” (Rawls, 1999, p. 9) and proposed fairness as a measure of justice. Recently, a lively debate has arisen around the fairness of new algorithmic technologies.

Algorithms are increasingly being integrated into decision-making frameworks, and there is growing evidence that these algorithms may perpetuate biases against certain groups of people. A prominent example is the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) software, designed to assess the risk of recidivism and inform pre-trial detention decisions (Larson et al., 2016). In the context of the United States, where there is a well-documented history of racialized policing (Hinton and Cook, 2021; Muller and Schrage, 2014; Western and Muller, 2013), concerns have been raised about the potential for racial disparities in the software’s predictions. At first glance, the algorithm seemed equitable: it detained white and black individuals at similar rates. However, a closer examination revealed significant bias (Chouldechova, 2016). The algorithm exhibited higher false positive rates for black offenders, incorrectly predicting that they would engage in future criminal activity more often than their white counterparts. As a result, black offenders were disproportionately subjected to wrongful pre-trial detention.

Algorithmic fairness challenges extend across various domains, including but not limited to, job advertising (Lambrecht and Tucker, 2018), childhood welfare systems (Chouldechova et al., 2018), credit approval processes (Kozodoi et al., 2022), recommender engines (Schnabel et al., 2016), and facial recognition technologies (Russakovsky et al., 2015). Research in this area is multifaceted, exploring diverse topics that range from methodological assessments of machine learning fairness algorithms — such as regression (Berk, Heidari, Jabbari, Joseph, et al., 2017), principal component analysis (Samadi et al., 2018), and classification (Hardt et al., 2016) – to conceptual frameworks for fairness, including treatment equality (Berk, Heidari, Jabbari, Kearns, et al., 2017) and equal opportunity (Hardt et al., 2016). This body of work also examines the goals of fairness, whether oriented towards individual fairness or group fairness (Dwork et al., 2012). Additionally, a critical stream of research has illuminated the inherent limitations in achieving fairness, demonstrating both the impossibility of satisfying multiple fairness criteria concurrently (Pleiss et al., 2017) and the potential costs associated with specific fairness-oriented decisions (Corbett-Davies et al., 2017).

To assess algorithmic fairness, researchers have proposed various definitions and metrics. General definitions, like the ”absence of prejudice toward any individual or group” (Mehrabi et al., 2022, p. 2), are complemented by precise criteria that account for outcomes and similarities across groups or individuals. Goals may differ: individual fairness aims for equal treatment of similar individuals (Dwork et al., 2012), while group fairness seeks equal treatment for entire groups, like racial categories (Kleinberg et al., 2016). For instance, financial risk algorithms often prioritize individual fairness, being blind to attributes like race (J. Chen et al., 2019; X. Chen et al., 2019). Such algorithms may appear individually fair but can violate group fairness, especially when risk factors disproportionately affect certain racial groups. This poses a trade-off, as achieving group fairness by altering approval thresholds would compromise individual fairness (Corbett-Davies et al., 2017).

To evaluate fairness, researchers employ fairness metrics, which consist of an outcome measure and a similarity measure (Mehrotra et al., 2022). Outcome measures quantify a specific outcome, like the average income across racial groups. Similarity measures, such as the statistical rate, assess how these outcomes vary between groups (Pessach and Shmueli, 2022). The statistical rate divides the minimum by the maximum observed outcome, generating a fairness score between 0 and 1—where 0 indicates perfect inequality and 1 indicates perfect equality. This approach provides a straightforward way to gauge whether an outcome satisfies the fairness definition’s similarity conditions.[2]

To illustrate the practical application of fairness metrics, let’s consider a scenario where 75% of individuals in Group A receive housing vouchers, compared to only 50% in Group B. To assess the fairness of this distribution, one could adopt an individual-level definition, stipulating that everyone should have an equal chance of receiving a voucher, irrespective of their group affiliation (Verma and Rubin, 2018). Utilizing the statistical rate as the similarity measure yields a fairness metric of 0.667. This suggests that the likelihood of receiving a housing voucher is only 67% as high for members of Group B as for those in Group A. Such a result indicates inequity, as it falls short of the ideal fairness metric of 1, thereby failing to satisfy the chosen definition of fairness.

[2] Our focus is on assessing similarity rather than using lower bounds like the 80% rule (Foulds et al., 2020), which falls outside this project’s scope.

The choice of measurement strategies can significantly impact the conclusions drawn in fairness assessments (Coulter, 2019; Cowell, 2011). For instance, some similarity measures focus solely on the most disparate outcomes, ignoring all other data points, and thereby produce a fairness metric that highlights only the most extreme inequalities. Additionally, the specific outcome selected for measurement is crucial to the analysis. Different outcomes, such as wealth versus income, can yield divergent results regarding inequality (Narayanan, 2018; Zafar et al., 2017).

1.2 Fairness in Diffusion

The diffusion process serves as one mechanism for distributing resources within a society, impacting not only the allocation of material goods but also the dissemination of information and the spread of contagion. These processes frequently manifest as emergent phenomena, resulting from the collective interactions of numerous individuals rather than the deliberate actions or attributes of any single participant (Rogers, 2003). Consequently, diffusion processes can inadvertently contribute to the creation, reinforcement, or perpetuation of social inequities (DiMaggio and Garip, 2011; 2012). The perception of these processes as either fair or unfair is contingent upon the particular framework of fairness that is employed for evaluation (Stoica, 2020; Venkatasubramanian et al., 2021).

A wealth of research underscores how network-based processes can exacerbate inequalities among groups distinguished by attributes like race or gender. These disparities manifest in various contexts, including the dissemination of job-related information (Ioannides and Loury, 2004; Pager and Pedulla, 2015; Pedulla and Pager, 2019; Petersen et al., 2000), the distribution of power within workplaces (Ibarra, 1992; Kalev, 2009; Yang and Aldrich, 2014), public health initiatives focused on disease prevention (Campos-Castillo and Laestadius, 2020; Gauthier et al., 2020), and exposure to predatory advertising (Barbeau et al., 2005; Graves, 2003; Wyly et al., 2006), among other areas.

Fundamental Cause Theory provides a valuable framework for comprehending how systemic disparities in opportunities are distributed across demographic groups such as race, gender, and socioeconomic status (Harder and Sumerau, 2018; Link and Phelan, 1995; 2010; Lutfey and Freese, 2005; Phelan and Link, 2015). While the theory is most frequently applied to health inequalities, it also offers insights into issues of socioeconomic status (SES) and discrimination in various domains. Phelan and Link, for instance, explicitly identify discrimination as an influential factor in SES, which in turn shapes assortative patterns in education, occupation, income, and residential location (Marsden, 1987; 1988; McPherson et al., 2001; Phelan and Link, 2015). Discrimination further erodes access to advantageous social connections, power, and prestige through mechanisms like stereotyping and implicit bias (2015, p. 318). In essence, an individual's group affiliation can influence their structural position within a network and, due to the phenomenon of homophily (in which similarity breeds connections), limit their ability to access and disseminate information within a constrained network environment (Aral and Van Alstyne, 2011; Hansen, 1999; Reagans and McEvily, 2003).

In diffusion processes, elements such as information, beliefs, attitudes, and contagions traverse from one individual to another at varying speeds, influenced by factors such as transmission mechanisms, individual network positions, and the overall topology of the network. These variables can differ between groups while showing internal consistency within groups. Some methods of transmission are intrinsically faster given network topology, but factors like cost, familiarity, and individual attitudes can affect the efficiency of information flow in groups where these factors are constraints (Centola, 2015; DiPrete et al., 2011; Levin and Cross, 2004). Individuals with high influence are adept at disseminating simple information but may struggle with complex topics due to the limitations of managing numerous social ties (Cross and Sproull, 2004; Hansen, 1999; Reagans and McEvily, 2003). Those with high betweenness centrality can efficiently circulate information to distant clusters in the network but may also strategically withhold it for reasons such as tactical advantage, a lack of mutual understanding, or even discrimination (Burt, 2007; Padgett and Ansell, 1993; Pedulla and Pager, 2019; Valente and Fujimoto, 2010). Groups characterized by homophily may reinforce information internally but are less likely to receive it from external groups (Centola and Macy, 2007). Thus, network characteristics within different groups can shape the balance of information spread in a multitude of ways.

Fairness in diffusion processes is an emerging focus within the field of computational and algorithmic fairness. A growing body of simulation-based research highlights various factors that can exacerbate outcome disparities across different groups. These factors include, but are not limited to: increased levels of homophily in networks compared to random graphs (Stoica, 2020; Stoica et al., 2020), the imposition of time constraints on diffusion processes (Ali et al., 2019), unequal degree distributions among groups (Stoica and Chaintreau, 2019), and the potential for advantageous network positions to perpetuate positive outcomes within certain groups, thereby widening existing inequities (Fish et al., 2019; Venkatasubramanian et al., 2021).

Notably, much of the existing research is oriented towards large-scale digital social networks, often employing simplified models wherein nodes lack specific social attributes and diffusion is presumed to be simple (Rahmattalabi et al., 2020; Tsang et al., 2019). In more realistic contexts, however, social characteristics — including differing beliefs, attitudes, or values — may substantially influence the reception and transmission of information or resources (Valente and Pitts, 2017; Watts and Dodds, 2007). Additionally, network topologies may vary across groups in ways that are not easily formalized but are nonetheless impactful.

1.3 Seeding Strategies as Network Intervention

Interventions aimed at shaping network diffusion often emphasize strategic seeding techniques. While information can propagate through a network spontaneously, as is often the case with rumors (Banerjee et al., 2014; Godes and Mayzlin, 2004), it is also possible to manipulate the diffusion process by deliberately introducing information to a targeted subset of individuals within the network. In some of the seminal empirical work in this field, researchers examined how the adoption of high-yield hybrid corn seeds disseminated among a community of Iowa farmers (Ryan and Gross, 1950; Valente and Rogers, 1995). Perhaps because of its roots in agricultural science, the initial subset of individuals exposed to treatment are referred to as “seeds.”

The classic approach to seed selection in network interventions, which remains particularly relevant for hard-to-reach populations, involves administering surveys within the target demographic. Respondents are prompted to identify their social connections based on a specific relational context believed to facilitate the diffusion process. For instance, in an influential study concerning the diffusion of a new pharmaceutical among physicians, participants were asked, ”To whom do you most often turn for advice and information?” (Coleman et al., 1957, p. 254). These surveys generate ego-centric data, serving to identify the most frequent nominees as the most influential ’seeds’ within the network. In essence, this nomination-based method gauges individuals by their degree centrality — where ’degree’ refers to a type of edge or relationship the researcher deems crucial to the adoption mechanism. This approach retains its utility in contemporary interventions involving resources that require elevated levels of trust, mutual understanding, and social reinforcement. Examples of applications in this category include health programs (Kim et al., 2015; Lungeanu et al., 2021), suicide prevention (Pickering et al., 2018), anti-smoking and e-cigarette campaigns (Chu et al., 2021; Cobb et al., 2016; Valente et al., 2003), HIV prevention (Amirkhanian et al., 2005; Kelly et al., 2000; Kelly et al., 1997; Kelly et al., 2006; Yadav et al., 2018), substance abuse countermeasures (Valente et al., 2007), and sexual violence prevention (Edwards et al., 2022).

Much existing theory surrounding targeted interventions in diffusion processes is designed for contexts where traditional survey methods may not be applicable or efficient. Such settings frequently include online social networks and other large-scale social systems where the logistics of survey administration become impractical. In these cases, seed selection often relies on various centrality measures to identify influential nodes. Degree centrality, which quantifies the number of social ties a node has, serves as an intuitive initial heuristic. Nodes with a higher degree of centrality are not only more likely to be information-rich due to their extensive connections but also often possess elevated social status (Goldenberg et al., 2009). Theoretically, their enhanced knowledge and status may endow them with greater influence within the network. Degree centrality is a commonly employed heuristic in a wide array of studies; examples cited here, such as the diffusion of pharmaceuticals (Coleman et al., 1966; Iyengar et al., 2015) and viral marketing in scale-free networks (Hanaki et al., 2007; Hinz et al., 2011; Van den Bulte and Joshi, 2007), represent a small subset of the research in this area.

While degree centrality often serves as a simple heuristic for seed selection, it may falter in certain network configurations—specifically, in highly clustered networks where seeds could struggle to facilitate information flow to remote clusters. In network topologies characterized by distinct clusters connected by long-range ties, nodes with high betweenness centrality — those serving as critical connectors or bridges between clusters — may prove to be more effective seeds for diffusion than nodes with high degree centrality (Freeman, 1977). This approach has been employed in various contexts, from suppressing epidemics (Schneider et al., 2011) to disseminating agricultural information (Beaman and Dillon, 2018), and in studies of disease spread (Saraswathi et al., 2020; Valente and Fujimoto, 2010).

Seed selection strategies often deploy a hybrid approach, combining elements of both degree and bridging properties. Metrics such as shortest path or closeness centrality incorporate both the spanning capability and the informational bandwidth of nodes, providing more comprehensive seeding heuristics (Aral and Van Alstyne, 2011; Freeman et al., 1979). Other advanced network metrics like K-core, percolation, and influence maximization aim to optimize the seeds' reach across the network, maximizing the extent of diffusion (Kempe et al., 2003; Kitsak et al., 2010; Morone and Makse, 2015).

The selection of a heuristic for seed identification is, of course, contingent upon the specific contextual factors of the diffusion process. In relatively simple networks such as scale-free networks, opinion leaders can often be pinpointed using less complex metrics like degree centrality or eigenvector centrality (Burt, 1999; Iyengar et al., 2011; Valente and Davis, 1999). When the likelihood of node-to-node transmission is particularly low, degree centrality effectively approximates eigenvector centrality and influence maximization (Kempe et al., 2003). However, in scenarios where transmission likelihood is high or varies across the network, more sophisticated algorithms like influence maximization may be more effective.

The seed selection methodologies discussed thus far present certain limitations when applied as interventions. Firstly, advanced strategies like influence maximization may yield more efficient diffusion but come at the cost of requiring extensive network data, in contrast to simpler heuristics such as degree centrality. Secondly, a foundational assumption underpinning these seeding strategies, from rudimentary heuristics like degree centrality to more advanced algorithms like influence maximization, is the principle of simple contagion:

the notion that resources or information can be transmitted to one individual from one individual (Granovetter, 1973; Watts, 1999). This assumption is often modeled analogously to the spread of infectious diseases. However, many diffusion processes more closely resemble knowledge transfer or resource adoption scenarios, which may involve non-trivial costs or require a certain level of mutual understanding, trust, or social reinforcement (Aral and Van Alstyne, 2011; Centola, 2015; Hansen, 1999; Levin and Cross, 2004). In such contexts, where the diffusion process more closely aligns with what is known as complex contagion, resources typically propagate through multiple neighboring nodes before successfully influencing a specific node (Centola and Macy, 2007). For interventions targeting complex contagions, it is beneficial to employ a seeding strategy focused on nodes with high ”complex centrality” (Guilbeault and Centola, 2021), a metric designed to identify nodes connected via sufficiently broad pathways to others within the network.

Network interventions, as we’ve discussed up to this point, often overlook the potential for the network feature driving the intervention to be strongly linked with a pre-existing advantaged group. This oversight could inadvertently lead to biased or unrepresentative seed sets, as well as unequal outcomes downstream. To illustrate, let’s imagine we are orchestrating a network intervention aimed at spreading COVID-19 prevention practices among high school students. We employ a basic heuristic for seeding, selecting 10 students based on the highest degree centrality. But what if all these seeds are male? Or exclusively white? Such a selection not only skews representation of early adopters but also risks marginalizing female or non-white students, especially if homophily is present in the high school network, as is often the case (McPherson et al., 2001). As we transition to the next section, we will delve into the concept of group fairness in relation to diffusion processes, addressing these very concerns.

1.4 Considering the Fairness of Diffusion: Intervention and Outcome

There are two primary points where inequities of network diffusion processes can be accounted for: at the point of intervention and at the culmination or outcome of the diffusion process. Historically, research on interventions for diffusion processes has prioritized maximizing efficiency (M. Li et al., 2017; Y. Li et al., 2018; Valente and Vega Yon, 2020). Interventions of this type target individuals more likely to possess network features conducive to information spread. Opinion leaders, for instance, can influence the behavior and opinions of others (Valente and Pumpuang, 2007). Influential individuals often possess a mix of strong and weak ties, which allows them to bridge disparate clusters in a network and thus become critical points for information diffusion (Granovetter, 1973). Those that can bridge structural holes can access diverse information and control its flow (Burt, 1995). Influential individuals can also exist at the core of the network and enhance their influence by being well-connected to other well-connected individuals (Kitsak et al., 2010). They often find themselves in roles that help optimize modularity (Nematzadeh et al., 2014).

In certain contexts, it's common for influential individuals to be situated in similar demographic groups, often reflecting shared attributes like race or gender. Organizational roles that determine network influence based on hierarchical structures frequently display uneven representation across different races and genders (Acker, 1990; Elliott and Smith, 2004; Maume Jr., 1999; Zweigenhaft and Domhoff, 1999). Minority groups, which commonly encompass women and racial minorities, often have limited influence within organizations (Kanter, 1993). Ibarra suggests that women's reduced access to organizational power stems from the nature of their networks where the tendency for individuals to connect with those similar to themselves serves as a barrier to entry (Ibarra, 1992).

Influential individuals often form connections based on mutual experiences, educational backgrounds, or professional paths (Zuckerman and Jost, 2001). These shared characteristics shape primary discussion networks (Marsden, 1987; 1988; McPherson et al., 2006), which frequently exhibit patterns based on race and gender (McPherson et al., 2001). More frequent interaction between influential individuals emerges from having shared resources, objectives, and ways of communicating (Eagle et al., 2009; Freeman, 1978; Padgett and Ansell, 1993). Consequently, it is common for influential individuals to belong to the same privileged demographic groups.

As a result, initiating a diffusion process with the most influential individuals — without accounting for demographic groups — can inadvertently favor privileged demographics due to the prevalence of homophily among these individuals. This can lead to noticeable disparities in how information spreads across groups. The inclination of influential individuals to connect predominantly with those from their own group means that information introduced within such a group is likely to be disseminated efficiently among its members before reaching individuals in outside groups (Reagans and McEvily, 2003). Thus, less privileged groups tend to access this information later and through longer network ties that have less capacity to channel large amounts of information (Aral and Van Alstyne, 2011). Such delays and dilutions exacerbate the existing challenges faced by less privileged groups, further widening network disparities with respect to privileged groups.

In this context, the concept of group fairness emerges as a critical point of concern. We can assess the fairness of the initial intervention in terms of group representation — specifically, on what basis are individuals selected as seeds? Additionally, we need to evaluate the fairness of the diffusion outcome among different groups — namely, was information dissemination equitable across all groups? A fair intervention ensures that no single group possesses an undue advantage in disseminating information across its own members. Conversely, a fair outcome is characterized by the absence of any group receiving a disproportionate share of the information.

Selecting seeds equitably appears straightforward, yet ensuring group fairness at this stage does not guarantee an equitable diffusion outcome. Fundamentally, the diffusion process represents a complex system (Easley and Kleinberg, 2010). At its core, a complex system is composed of many parts that interact in non-simple ways. The emergent properties of the system, which arise from these interactions, cannot be easily predicted by examining the parts separately. A key characteristic of complex systems is non-linearity. In the context of information diffusion, this means that the relationship between the input (e.g., initial seed nodes) and the output (e.g., total nodes reached) is unlikely to be proportional (Lehmann and Ahn, 2019; Porter and Gleeson, 2016).

Within a single social network, two distinct groups can be conceptualized as two subgraphs. Not only might the nodes in each subgraph vary in terms of characteristics and mechanisms for transmitting information, but the topologies of these subgraphs can also differ. Differing mechanism of transmission change the dynamics and efficiency with which information can spread within a group. For instance, one group might have a higher threshold of a adoption. In this case, the early stages of diffusion through the group may be slow, but could become more efficient once a critical mass is reached (Centola et al., 2007; Granovetter, 1978; Valente, 1996). The distribution of influence among key individuals in each subgraph might differ, leading to unique dynamics within groups (Stoica et al., 2020). Additionally, each subgraph could possess distinct distributions of particular network properties, potentially enhancing or limiting the spread of information, contingent on the interplay between individual characteristics, transmission mechanisms, and network topology (Porter and Gleeson, 2016). The links connecting the subgraphs introduce an additional layer of intricacy, affecting the flow of information between groups.

2 Data and Methods

For our analysis, we draw upon network data provided by Ferrali (Ferrali et al., 2020), which captures real-world information diffusion processes across 16 independent Ugandan villages. In these communities, a new Political Communication Technology (PCT) was introduced. The villages were selected based on the adoption rates of the technology; half exhibited high uptake rates, while the other half showed low uptake rates. This technology enabled users to disseminate messages on local and regional issues, including but not limited to health, education, water supply, and infrastructure.

Utilizing the independent village data allows us to approximate diffusion models under conditions that are more realistic than those in fully simulated experiments. This is particularly important as the technology at the center of the diffusion process is designed to foster political participation. Given that political participation often varies between different social groups, understanding the dynamics of this diffusion is an important application of group fairness to equality in political participation.

In the initial stage of Ferrali et al. study, a select group of villagers were invited to a meeting where the technology was demonstrated. These attendees subsequently became the self-selected initial nodes—or ”seeds”—of the diffusion processes.[3] Two years later, researchers collected data on awareness of the technology and adoption within each village. The data included four types of directed ties among villagers: family, friend, lender, and problem-solver, along with individual-level demographic data.

Our focus is on gender-based differences in information diffusion about the technology.[4] In each village, women were less likely than men to be informed about the PCT, indicating disparate impacts across villages based on gender. While our analysis centers on gender, it is important to note that other social categories could also provide valuable insights.

Our study encompasses 14 out of the original 16 villages, as two were excluded due to incomplete gender data. We also narrowed our focus to two types of directed edges: family and friend ties, primarily because lender and problem-solver ties were made redundant by family and friend ties, and tie-type was not a parameter of the simulation model.

The networks under study comprise 3,184 nodes in total, representing approximately 82% of the adult population across the villages, and 41,495 edges. This rich dataset enables us to examine individual network positions, overall network topology, and group membership categorized by gender.

2.1 Network Centrality Measures

In our empirical investigation, we employ an array of seeding strategies to explore interventions in information diffusion processes. These seeding strategies are aimed at identifying opinion leaders or influential nodes based on network structure. Each strategy is one traditionally thought of as conducive to a specific type of network topology: degree centrality is effective in homogeneous or well-mixed networks (Newman, 2018); betweenness centrality is ideal for networks with distinct clusters connected by bridging nodes (Freeman, 1977; Valente and Fujimoto, 2010); closeness centrality is suitable for small-world networks (Watts, 1999); eigenvector centrality works well in networks where influence is distributed through key nodes (Bonacich, 1987); complex centrality is ideal for networks that require a high degree of mutual reinforcement for information spread (Centola and

[3] It is important to note that we lack information on the criteria used for selecting villagers for the initial meeting. However, we can confirm that the selection did not involve any maximizing network strategies. The network survey was conducted two years post-technology introduction.

[4] Our focus is on information spread because it is commonly treated as simple contagion where the mechanism of transmission

Macy, 2007); influence maximization is effective in scale-free networks or those with heterogeneous degree distributions (Kempe et al., 2003); percolation centrality is beneficial in networks vulnerable to cascading failures (Morone and Makse, 2015); and k-core centrality is applicable in hierarchical or modular networks (Kitsak et al., 2010).

Degree Centrality: is quantified as the number of connections or ties that a node maintains within a network. As one of the most straightforward centrality measures, degree centrality operates on the premise that nodes with higher degree centrality are optimally positioned to disseminate information to a larger subset of individuals within the network.

Betweenness Centrality: quantifies the frequency with which a node serves as a bridge along the shortest path between two other nodes in the network (Freeman, 1977). As a well-established centrality measure, betweenness centrality measures the potential influence of an individual within the network based on their recurring role in connecting disparate clusters or nodes (Padgett and Ansell, 1993). Nodes with elevated betweenness centrality often act as pivotal conduits, facilitating the entry and dissemination of information across distinct regions of the network.

is often assumed to be uniform across groups

Closeness Centrality: quantifies the average distance from a specific node to all other nodes within the network, typically calculated by taking the sum of the lengths of the shortest paths between the node in question and every other node (Freeman et al., 1979). Often, this sum is inverted to create a measure where higher values indicate greater closeness—and by extension, potential influence—within the network. Nodes with elevated closeness centrality are strategically positioned to disseminate information, as they maintain comparatively shorter paths to all other nodes. Consequently, such nodes are equipped with the capability to exert significant influence across the network, owing to their relative proximity to all other nodes.

Eigenvector Centrality: expands on the concept of degree centrality by not only considering the number of direct ties a node has but also evaluating the centrality of those connected nodes (Bonacich, 1987). Unlike basic degree centrality, which may inadvertently give high centrality to a node connected only to less central nodes, eigenvector centrality assigns weighted significance to nodes that are connected to other highly central nodes. This measure is mathematically derived from the eigenvector associated with the network's adjacency matrix, providing each node with a score that reflects both its direct connections and the centrality of its neighbors. This nuanced approach enables a more comprehensive assessment of a node's influence within the network.

Complex Centrality: is predicated on a distinct conceptual model of diffusion processes, diverging from the notion of simple contagions where a single exposure suffices for transmission (Centola et al., 2007; Centola and Macy, 2007). In contrast, complex centrality is designed to quantify the dynamics of complex contagions, where multiple exposures from different nodes may be required for successful transmission (Guilbeault and Centola, 2021). This approach is particularly applicable to the dissemination of norms, beliefs, and intricate information, which often necessitate trust, mutual understanding or reinforcement from multiple social contacts. The metric identifies the prevalence of sufficiently broad bridges in the network that facilitate the complex contagion process beyond a specified threshold. In our empirical analysis, we examine the strategy with contagion thresholds of $T = \{2,4,6\}$.

Influence Maximization: employs an algorithmic approach to identify a subset of nodes that optimally maximizes the reach of influence within a network. Starting with a predetermined interaction probability between node i and its neighbor node u, the algorithm sets this probability uniformly across all edges in the network (Kempe et al., 2003). Activation of a particular node j occurs only if it is part of a "live-edge path" — a sequence of activated edges — that connects it to another active node within the network. The expected influence spread from one node to another is calculated based on the likelihood of a live-edge path existing between them. Utilizing a greedy selection method, the algorithm evaluates seeds according to their potential influence spread, measured by the number of live-edge paths under a fixed probability p. Seeds are iteratively chosen and removed based on this metric until a set of the desired size is obtained. The end result is an optimized set of seed nodes that maximize the potential for influence across the network. In our study, we test the effectiveness of the influence maximization algorithm using four different transmission probabilities $p = \{0.01, 0.10, 0.25, 0.50\}$.

Percolation: quantifies the likelihood that a connective path exists between two nodes within the network. This metric is often synonymous with the term 'collective influence' (Morone and Makse, 2015). Specifically, the percolation value is calculated as the product of a node's reduced degree and the aggregated reduced degrees of all nodes situated at a distance d from the target node. A distance parameter of $d = 3$ was selected for analysis, aligning closely with the average inter-node distance observed across the surveyed villages.

K-Core: quantifies the maximum subgraph in which each vertex exhibits a minimum degree of k. The 'coreness' of a vertex is determined if it is part of the k-core but not of the $(k + 1)$-core. Calculation of the k-core metric begins with an assessment of each node's degree. Subsequently, the k-core algorithm iteratively excises all nodes possessing a degree less than k, repeating the process until the remaining nodes all have a degree of at least k. This maximal subgraph then serves as the k-core network measure. Employed for identifying core nodes within a network, k-core can be a valuable strategy for seed selection (Kitsak et al., 2010).

2.2 Na¨ıve Seeding of Diffusion Processes

In the set of cases examined in this study, Ugandan villagers from 14 independent villages were invited to a meeting (in each village) about the new political communication technology (PCT). Those who attended were designated as the self-selected seeds for spreading information about the PCT for each village.

Group fairness in this context implies that both male and female villagers are treated in a roughly equitable manner. While interpretations of "equitable" can differ, a baseline definition for fair treatment — or the fair selection of seeds — is the stipulation that the probability of a villager being selected as a seed is uniform regardless of being male or female. In simple terms, this means that each demographic group should have proportional representation among the seed set.

Network interventions often overlook Group Fairness when focusing on the maximization of influence through seed selection. By initiating the diffusion process with the most influential individuals in a social network, there's a risk of inadvertently choosing a disproportionate number of members from a privileged group.

Figure 1 presents the distribution of seeds among male and female villagers across all 14 villages for the self-selection seeding and a set of maximizing strategies. The x-axis indicates the proportion of the female seed set relative to the female population, while the y-axis indicates the proportion of the male seed set relative to the male population (in each village). The dashed line following the equation $y = x$ signifies a perfect parity between male and female seed sets in relation to their respective populations. Points situated below this line represent interventions skewed toward female villagers, while those above signify interventions skewed toward male villagers.

On the right side of Figure 1, each network maximizing strategy (defined in subsection 2.1) is paired with the corresponding proportion of female and male seeds relative to their population sizes. These maximizing strategies represent seed sets in which the corresponding network measure has been optimized. The results are averages of seed selections across all 14 villages. The data points in the figure should form a relatively straight line; however, they've been adjusted for easier readability, with the exact values provided. As a side note, ”self-selection” denotes the seed set from the actual diffusion processes that occurred. This strategy is termed ”self-selection” as villagers voluntarily participated in PCT information sessions.

Figure 1 demonstrates that each maximizing strategy would not have an equitable representation of both male and female villagers. Most of the seeding strategies, including self-selection, lie above the equality line, indicating an over-representation of male villagers as seeds. The two notable exceptions are complex centrality with thresholds of 2 and 4, which, when maximized, would disproportionately select female villagers. In an interesting contrast, complex centrality with a threshold of 6 hovers near the equality line but would technically contain an over-representation of male villagers. Eigenvector centrality emerges as the most male-biased seeding strategy, followed by closeness, influence maximization, and degree. Under a simple contagion assumption for information dissemination, this could favor male villagers. However, when we assume a moderately complex contagion — where individuals require either two or four informed neighbors to become informed — female villagers would be disproportionately selected as seeds. Segments of female villagers in each village must possesses more paths of sufficient width, giving them higher complex centrality at these thresholds. Yet, when the threshold rises to 6, male villagers become disproportionately selected.

While these results underscore disparities in group representation over various maximizing strategies, they also raise questions about how such imbalances impact the outcome of information diffusion processes.

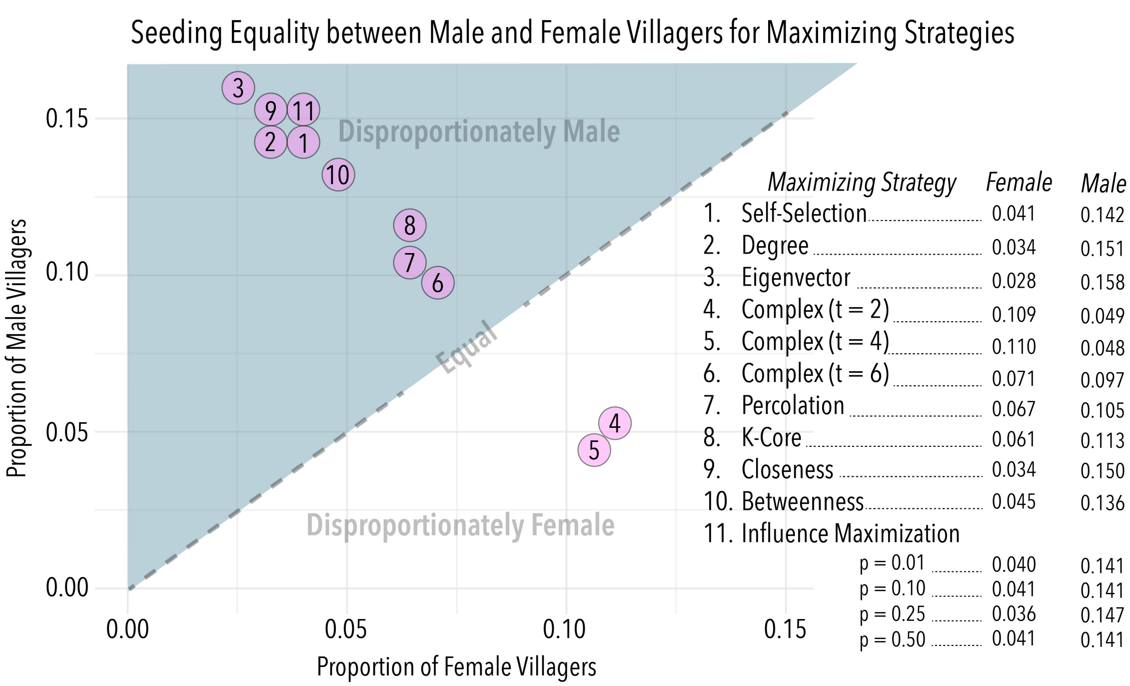

Figure 1: Equality of Naive Seeding

To explore this, the subsequent section introduces the diffusion model used in this study. With this model, we can assess the impact of information diffusion on male and female villagers, and explore the relationship between the fairness of initial seeding strategies and the equitable dissemination of information.

2.3 Approximating the Empirical Diffusion Process

To examine the effect of different seeding interventions, we construct a model that approximates the actual diffusion processes within each village using a simulated method of moments (SMM) approach (Evans, 2023). The SMM is a robust, non-parametric technique adept at estimating parameters from intricate probability distributions. We acheive this by first specifying a simple diffusion model that creates a decision-making framework for our settings. To approximate the actual diffusion processes, we need to identify the likelihood that a villager will share information about the PCT with their neighbors under specific conditions. Using SMM, we locate this likelihood by searching across a large set of parameter combinations and identifying the optimal set through calibration simulations. Essentially, we repeatedly simulate the diffusion process using various parameter values until we reproduce a process that closely aligns with observed empirical data. The degree of this alignment is gauged using 'moments', which are statistical snapshots of the diffusion process — for example a moment could be, the total degree of individuals who were informed. By evaluating the differences in moment values between our simulated and empirical processes, we identify the parameter set that best approximates the actual diffusion process. With this model, we can then simulate a range of alternative network interventions.

The Simulated Method of Moments (SMM) is an extension of the Generalized Method of Moments (GMM). One of its main advantages over GMM is its adaptability, allowing for the incorporation of intricate models and handling a wider variety of data. SMM facilitates the fitting of our model without necessitating stringent assumptions about the theoretical distribution. While there are alternative methodologies to SMM, each carries its own set of constraints, which underscores the advantages of an SMM approach. Although we could consider methods such as maximum likelihood estimation, Bayesian estimation, or ordinary least squares, they encounter challenges with intricate models, often demand robust prior assumptions, and can yield biased outcomes if data is not normally distributed.

2.3.1 Specifying the Diffusion Model

We utilize a simple diffusion model, as outlined by (Banerjee et al., 2013). The model operates as follows:

1. A set of seeds is identified, and they are informed about the technology.

2. Each seed then decides to adopt the technology based on an individual probability.

3. If a seed adopts, they convey the information to each of their neighbors independently with a probability P_A.

4. Conversely, if they opt against adoption, they share the information with each neighbor with a probability P_{NA}.

5. Steps 2 thru 4 are subsequently re-initiated for the newly informed individuals and continues across six time frames.

For the second step, villagers' adoption probabilities are derived using logistic regression likelihood estimates. These probabilities are computed based on individual-level attributes from the survey data (Ferrali et al., 2020). Table 1 provides four distinct models. Specifically, the fourth column contains a model narrowed down to significant covariates: degree (cont), age (cont), edu (0,1), hasPhone (0,1), and female (0,1), where adoption (0, 1) serves as the dependent variable.[5] These estimates were formulated using a pooled sample of 80% of villagers across all villages, with standard errors clustered at the village level, ensuring convergence in estimation. After receiving information about the PCT, each villager must decide on adoption. The model emulates this decision by assessing if a villager's predicted adoption likelihood, derived from the fourth model in Table 1, surpasses a randomly generated value between 0 and 1. A villager opts for PCT adoption when their predicted probability exceeds this random value.

[5] Degree refers to a villager's total number of neighbors. Education is a binary variable that is 1 if the villager attained at least secondary education.

Logistic Regression Estimates

	Dependent variable:			
	adopt			
	(1)	(2)	(3)	(4)
degree	0.063***	0.072***	0.077***	0.076***
	(0.018)	(0.017)	(0.016)	(0.017)
age	−0.015	−0.018*	−0.015*	−0.015*
	(0.010)	(0.010)	(0.009)	(0.008)
edu	2.066***	1.962***	1.964***	1.988***
	(0.265)	(0.231)	(0.235)	(0.230)
female	−0.864***	−0.864***	−0.877***	−0.870***
	(0.299)	(0.303)	(0.309)	(0.307)
hasPhone	0.969***	1.012***	1.013***	1.034***
	(0.349)	(0.354)	(0.355)	(0.364)
income2	−0.078	−0.077	−0.082	
	(0.235)	(0.283)	(0.278)	
income3	0.345	0.377	0.382	
	(0.266)	(0.296)	(0.296)	
income4	0.138	0.100	0.108	
	(0.264)	(0.297)	(0.292)	
income5	0.712	0.561	0.553	
	(0.786)	(0.785)	(0.797)	
leader	0.323	0.332		
	(0.354)	(0.329)		
distMeeting	−0.061			
	(0.122)			
Constant	−4.923***	−5.038***	−5.132***	−5.053***
	(0.689)	(0.806)	(0.741)	(0.690)
Note:			*p<0.1; **p<0.05; ***p<0.01	

Table 1: Logistic Regression Models for Probability of Adoption

During the third and fourth steps, seeds undertake a probabilistic decision-making process to determine the transmission of information about the PCT to their adjacent nodes or neighbors. For each adjacent node, the decision to transmit information operates independently. Given a villager's choice to adopt the PCT, there exists a probability P_A associated with transmitting this information to each respective neighbor. In contrast, if the villager does not adopt, the transmission probability for each adjacent social tie is represented as P_{NA}. Within the context of our diffusion model, the probabilities P_A (probability contingent upon adoption) and P_{NA} (probability given non-adoption) are hyperparameters that must be calibrated. The calibration procedure for these values is detailed in subsubsection 2.3.2.

In step 5, steps 2 through 4 are iteratively executed for a total of six discrete time periods. The choice of six periods, while somewhat arbitrary, is due to the absence of granular, period-to-period data regarding the actual diffusion processes. Consequently, we lack the capability to tune period-specific outcomes between the modeled and empirical processes. Nonetheless, we are informed that diffusion occurred over a span of 18 months. Given this context, we deemed it appropriate to segment this timeline into three-month intervals, culminating in six distinct phases for the diffusion process.

2.3.2 Fitting the Model

To fit the model, we traverse a parameter space to identify our target parameters, P_A and P_{NA}. Specifically, P_A represents the probability that a villager, having adopted the PCT, will convey information to a designated neighbor. Conversely, P_{NA} represents the likelihood of information transmission when the villager has not adopted the PCT. These hyperparameters are tuned by conducting simulations over a discretized parameter space, subsequently assessing the distance between a given set of empirical moments and those produced within the simulated diffusion process.

An advantage of the Simulated Method of Moments (SMM) is its capacity to fit to an extensive set of moments, provided these moments a) are somewhat orthogonal to our target parameters and b) encapsulate essential relationships within the diffusion process. In this study, we calibrate diffusion models to each specific village and adopt a moment selection methodology to identify the most essential subset of moments.

We define the following 15 moments:

1. Proportion of isolated individuals informed.
2. Covariance between being informed and the proportion of informed second-degree neighbors.
3. Proportion of informed leaders.
4. Proportion of informed females.
5. Proportion of individuals informed despite having no participating neighbors.
6. Proportion of informed males.
7. Accumulated degree of informed individuals.
8. Mean proportion of informed neighbors.
9. Proportion of male adopters.
10. Proportion of female adopters.
11. Proportion of isolated individuals who adopted.
12. Covariance between adoption and the proportion of adoptive second-degree neighbors.
13. Proportion of adoptive leaders.
14. Accumulated degree of adopters.

15. Mean proportion of adoptive neighbors.

Each moment describes a statistical relationship among nodes, edges, and the transfer of information or the PCT. These moments represent an array of fundamental and more nuanced network outcomes that are directly or indirectly related to the spread of information (the process we intend to model). Importantly, not all 15 moments are employed for every model fit. To enhance the accuracy of each model, we incorporate a moment selection methodology. For each model, we simulate the diffusion process considering all 15 moments. Subsequently, we assess all potential combinations of a subset of 6 moments. After this assessment, we determine the 6-moment combination that most reduces the moment error function.[6]

During model calibration, each simulation batch evaluates a unique pair of P_A and P_{NA} values within the parameter space. This space is discretized, exhaustively exploring combinations of P_A and P_{NA} ranging from 0 to 1 in 0.02 intervals, resulting in a total of 2,601 parameter pairings. For every such pairing, we execute 100 simulations, thereafter computing the simulated moments as the mean across these simulations. This averaging procedure mitigates sensitivity concerns and enhances the precision of our parameter determination.

The quality of the model's approximation of the empirical process for each parameter value combination is assessed by calculating the error between the selected set of six simulated moments and the corresponding empirical moments. Calculation of the error is a procedure that requires three steps.

First, the error between the simulated moments and empirical moments is computed using a moment error function. For the purposes of this study, our chosen error function is the percentage difference between the simulated and empirical moments. By adopting the percentage difference, we standardize the moments to consistent units, thus ensuring that no particular moment is unnecessarily weighted. The moment error function can be expressed formally as[7]:

$$e(x, x|\theta) = \frac{m(x|\theta) - m(x)}{m(x)}$$

Here, θ signifies our parameter value of interest. The term $m(x|\theta)$ represents a vector encapsulating the average simulated moments derived from a series of 100 simulations, while $m(x)$ designates the matrix of empirical moments.[8]

[6] We chose 6 moments, after testing sets of size 4, 6, and 8, as the moment-set size that produced the best fit of the diffusion model to the empirical processes.

[7] Much of the following mathematical notation has been borrowed from (Evans, 2018) with the author's permission.

[8] The empirical moments matrix is a 15 x 14 matrix, where each column corresponds to the empirical moments of a specific village. The error computed is element-wise, as only one column of the empirical matrix is used (specific to the village).

In the subsequent step, we optimize the criterion function through a two-step procedure that involves applying specific weights to each moment to minimize the error between empirical and simulated moments. The objective is to identify an optimal weighting matrix characterized by the least asymptotic variance across the moments. Mathematically, this optimal weighting matrix is the inverse of the variance-covariance matrix of the moments, defined as:

$$W^{opt} = \Omega^{-1}(x,x|\theta)$$

Here, $\Omega(x,x|\theta)$ represents the variance-covariance matrix of the moment condition errors. To improve the accuracy of the model, it is important to minimize the weight for moments with high variance and, conversely, increase the weight for moments with low variance (e.g., moments that demonstrate more accuracy).

To incorporate optimal weighting within our criterion function, we transform the criterion function to the sum of squared errors using an identity matrix I, which is a 15 x 15 matrix characterized by ones along its diagonal. The first step of this optimal weighting is expressed as:

$$\theta_{SMM} = \theta : \min_{\theta} e(x,x|\theta)^T I e(x,x|\theta)$$

With this, we compute

$$\Omega = \frac{1}{N} e\left(x, x|\theta\right) e(x, x|\theta)^T.$$

The weighting matrix is then computed as the inverse: $W \equiv \Omega^{-1}$. Given that the matrix is non-singular, we compute the pseudoinverse via singular value decomposition.

Lastly, we recalibrate our parameter values employing the optimal weighting matrix:

$$\theta_{SMM} = \theta : \min_{\theta} e(x,x|\theta)^T W e(x,x|\theta)$$

The result is that we fit the model to the empirical diffusion processes by minimizing squared errors between the simulated moments and empirical moments with optimal weighting.

To ensure the accuracy of each model, we compare the cumulative information spread via the calibrated models using the actual seed sets against the actual diffusion processes for each village. The results of this model calibration are presented in Table 2. Our model closely approximates the empirical data on the information spread.

Calibration Results

	Adopted		Heard		Parameter Estimates	
Village	Calibration (avg.)	Actual	Calibration (avg.)	Actual	P_A	P_N
1	4.15	5	51.9	52	0.66	0.14
2	2.42	3	60.8	61	0.34	0.26
3	3.04	3	34.0	34	0.06	0.16
4	6.70	19	88.8	89	0.90	0.12
5	1.66	10	57.0	57	0.46	0.20
6	2.06	4	50.0	50	0.88	0.18
7	4.41	6	39.0	39	0.78	0.16
8	3.40	3	79.1	79	0.32	0.30
9	0.99	1	24.1	24	0.98	0.18
10	5.84	13	90.1	90	0.88	0.32
11	4.19	6	34.1	34	0.48	0.14
12	4.88	13	79.4	79	0.84	0.18
13	8.06	17	114.3	114	0.74	0.24
14	7.17	20	102.7	102	0.50	0.30

Table 2: Results of the Calibration of the Diffusion Model

2.4 Outcomes of Diffusion Processes with Maximizing Interventions

In subsection 2.2, we showed the representation of male and female villagers in seeding strategies that maximized a set of centrality measures (defined in subsection 2.1). We found that none of these seeding strategies, which did not consider equality, had equal representations of male and female villagers. Among the strategies, Self-selection, Degree, Eigenvector, Percolation, K-Core, Closeness, Betweeness, Influence Maximization, and Complex Centrality ($t = 6$) were disproportionately male. While Complex Centrality ($t = 2$) and ($t = 4$) were disproportionately female.

With the seeding strategies and the diffusion framework in context, we transition to question how these seeding strategies reverberate within the diffusion processes. Specifically, we are interested in the relationship between the intervention and the outcome, especially as it pertains to how information spreads across male and female villagers.

Figure 2 presents the outcomes of the maximizing strategies from subsection 2.2. Each maximizing strategy was simulated 1000 times in each village using the diffusion model (detailed in subsection 2.3). The outcomes in Figure 2 represent averages across the 14 villages. Each outcome corresponds to a distinct maximizing strategy. The y-axis illustrates the average fraction of informed male villagers, while the x-axis represents the same for female villagers. A dashed line, described by the equation $y = x$, signifies parity, indicating that information spread equally across both groups. Outcomes above the dashed line indicate that male villagers were more widely informed, whereas outcomes below denote female villagers were more

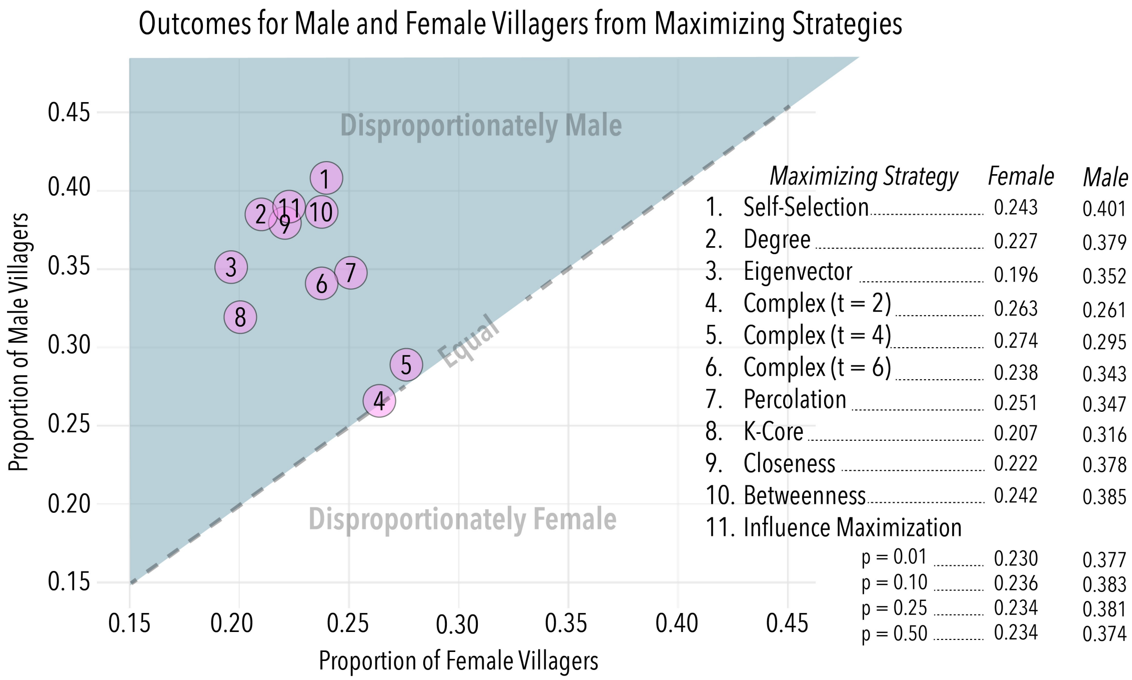

Figure 2: Equality of Outcome from Maximizing Seeds

informed. The table on the right side provides the exact proportions of informed male and female villagers for each strategy.

In Figure 2, we observe a distinct trend: the majority of maximization strategies lead to male villagers being disproportionately informed over female villagers. None of the strategies display a pronounced benefit for female villagers. However, two strategies (complex centrality: t = 2 and t = 4) yield outcomes that are approximately equal for both genders on average. When cross-referenced with Figure 1, maximizing complex centrality with thresholds of $t = 2$ and $t = 4$, we on average select seeds more from the female villager population across the 14 villages. This suggests that there are properties within these networks, or factors influencing villagers' likelihood to share information, that enable the male population to benefit equivalently from the diffusion process, even when they are less represented in the initial seed set. Thus, the seed distribution, determined by a particular network intervention, does not necessarily correspond to the ultimate distribution of informed individuals. However, all maximizing strategies that initially disproportionately selected male villagers also resulted in outcomes where male villagers were predominantly informed.

An important question is how do we interpret the magnitude of the disparities between male and female villagers in terms of the outcomes of the information diffusion processes? At first glance, the difference between an average of 25.1% of the female population being informed versus 34.3% of the male population (under maximized percolation) might appear marginal.

A practical metric to evaluate these outcomes is to determine the ratio of the lesser value to the greater value. This approach was demonstrated in our toy example presented in subsection 1.1. This measure, known as the *statistical rate*, draws a comparison between the two values. Its formal definition is as follows:

Given that M represents the set of male villagers measured and F stands for the set of female villagers measured, we can express:

$$SR = \frac{\min(r)}{\max(r)}, \quad r \in \{X_M, X_F\}, \qquad \text{(Statistical Rate)}$$

Here, X_M and X_F respectively denote the value measures associated with male and female villagers.

The statistical rate results in a real number bounded between 0 and 1. A value of 1 signifies absolute equality, while a value of 0 denotes complete inequality (e.g., a result of 0 for one of the groups). Intuitively, this rate can be interpreted as the fraction of the larger measure that the smaller measure represents, highlighting the relative scale of the outcome for the less-advantaged group in contrast to the favored group. The metric is useful in this context as we aim to contrast the proportions of a resource allocated to two distinct groups (Mehrotra et al., 2022).

Maximizing Strategy	*Female*	*Male*	*SR*
1. Self-Selection	0.243	0.401	0.605
2. Degree	0.227	0.379	0.599
3. Eigenvector	0.196	0.352	0.557
4. Complex (t = 2)	0.263	0.261	0.992
5. Complex (t = 4)	0.274	0.295	0.929
6. Complex (t = 6)	0.238	0.343	0.694
7. Percolation	0.251	0.347	0.723
8. K-Core	0.207	0.316	0.655
9. Closeness	0.222	0.378	0.587
10. Betweenness	0.242	0.385	0.629
11. Influence Maximization			
p = 0.01	0.230	0.377	0.611
p = 0.10	0.236	0.383	0.616
p = 0.25	0.234	0.381	0.614
p = 0.50	0.234	0.374	0.626

Table 3: Statistical Rates of Maximizing Strategy Outcomes

Table 3 extends the data from Figure 2 by introducing a column of statistical rates, representing the similarity of the average outcomes across villages between male and female villagers. These statistical rates provide a more straightforward interpretation of the relative outcomes across male and female villagers. We first observe that for complex centrality with thresholds at $t = 2$ and $t = 4$, the statistical rates are very near 1, indicating almost equal average outcomes. When evaluating the other maximizing strategies, the statistical rates provide a clearer view of the disparities between male and female villagers. When maximizing degree, eigenvector centrality, and closeness, the average outcomes for female villagers are less than 60% of what male villagers achieve. Apart from complex centrality, only the percolation strategy yields a statistical rate exceeding 0.7. This rate suggests that the outcome for female villagers was approximately 72.3% of the male villagers' outcome.

Having observed disparities across the majority of our network interventions, it is evident that maximizing different seeding strategies can produce imbalances both in initial seed selection and outcomes between male and female villagers. In the section that follows, we conceptualize new strategies aimed at achieving fairer outcomes across male and female villagers within the diffusion process.

2.5 Group Fairness for Seed Selection

Evaluating equitable outcomes between groups through the lens of group fairness necessitates an initial determination of what constitutes fairness. Within the purview of diffusion processes, fairness can be measured at two points: at the point of the intervention and at the realization of the outcome (refer to subsection 1.1).[9]We also must recognize that while seed selection remains a critical point of intervention for the diffusion process, realizing equitable outcomes is the primary goal of introducing group fairness constraints on the intervention. While equitable treatment or equity at the point of network intervention has important implications, realizing an equitable impact addresses deeper disparities between groups more effectively. The challenge of introducing group fairness constraints on network interventions is therefore identifying measures for seed selection that generate equitable outcomes in the culmination of the diffusion process.

In previous sections, we have hinted at an inherent definition for group fairness concerning outcomes. We conceptualize group fairness in outcomes as both male and female villagers possessing an identical likelihood of being informed. In this case, even if the female villager population is double that of male villagers, a 50% probability of being informed should, over a large number of trials, equate to 50% of both male and female villagers being informed. This can be defined formally as:

[9] We acknowledge the potential existence of alternate seeding strategies in diffusion processes, like the sequential initialization of multiple seed sets. However, such strategies fall beyond the ambit of this discussion.

$$P(i \in O : i \in M) = P(j \in O : j \in F),$$

where i and j denote villagers, M denotes the set of male villagers, F denotes the set of female villagers, and O corresponds to the set of informed villagers. This expression can be interpreted as "the likelihood that a villager i from set M is in the informed set, O, equals the likelihood that a villager j from set F is in the same informed set, O."[10]

Crafting a definition for fair seed selection poses more of a challenge. A straightforward definition for seed selection mirrors our definition for equitable outcomes between groups. Let us call this *simple seed fairness*, defining group fairness in seed selection to mean both male and female villagers have an identical likelihood of being chosen as seeds. This can be defined formally as:

$$P(i \in K : i \in M) = P(j \in K : j \in F), \qquad \text{(Simple Seed Fairness)}$$

where i and j denote villagers, M represents the male villager population set, F denotes the female villager population set, and K signifies the seed set. This equation can be interpreted as saying "the likelihood of a villager i from set M being in the seed set K is equal to the likelihood of a villager j from set F being in that same seed set, K."

Our definition for simple seed fairness, however, does not generate equitable outcomes between male and female villagers for our set of cases. Table 4 displays averages of the proportion of villagers from each group that were informed, as well as similarity measures of the outcomes (statistical rate) stemming from network interventions held to our definition for simple seed fairness. In these strategies, both male and female villagers received seed sets of equal sizes relative to their group size. Seeds for each group were chosen in descending order of their within-group rank for the network measure used in the seeding strategy. In other words, villagers were selected starting with the highest network measure within each group, with the number of seeds chosen for each group adhering to our definition for simple seed fairness. The values in the table represent averages across all 14 villages.

[10] This conceptualization of fairness can be seen as a nuanced adaptation of a classic fairness definition known as *statistical parity*. The distinction lies in our approach not specifically framing fairness within predictive tasks.

	Strategy	Female	Male	SR
1.	Degree	0.274	0.382	0.717
2.	Eigenvector	0.243	0.343	0.708
3.	Complex (t = 2)	0.246	0.338	0.728
4.	Complex (t = 4)	0.274	0.383	0.715
5.	Complex (t = 6)	0.271	0.373	0.726
6.	Percolation	0.281	0.379	0.741
7.	K-Core	0.225	0.312	0.721
8.	Closeness	0.267	0.368	0.725
9.	Betweenness	0.279	0.382	0.730
10.	Influence Maximization			
	p = 0.01	0.277	0.379	0.731
	p = 0.10	0.277	0.381	0.727
	p = 0.25	0.271	0.374	0.724
	p = 0.50	0.279	0.384	0.726

Table 4: Statistical Rates of Evenly Seeded Strategy Outcomes

By adhering to our simple seed fairness definition, we observe more equitable outcomes for the majority of network interventions relative to the previous maximized strategies for each network measure. Notably, the most equitable average outcome was realized when percolation was used as the network measure for selecting each group's seed sets. Even with this strategy, however, female villagers attained only 74.1% of the outcome realized by the male villagers. Moreover, the two complex centrality measures, which previously achieved near-perfect equity under maximization, now both have statistical rates of 0.728 (for $t = 2$) and 0.715 (for $t = 4$). This is an indication that other characteristics of the network or the population are influencing the results, highlighting the challenge of realizing fair outcomes solely through seed selection based on the simple seed fairness definition.

A significant limitation of the simple seed fairness definition lies in its inability to account for differences in network structure across groups. Allocating a proportional number of seeds to each group fails to address disparities in influence inherent to each group. Our observations, particularly when employing a naive maximization of our network measures (refer to subsection 2.2), revealed that male villagers exhibit more dominant distributions of the centrality measures used for these network interventions, barring the instances of complex centrality at $t = 2$ and $t = 4$.

These centrality metrics, elaborated in subsection 2.1, are chosen as seeding strategies primarily due to their frequent treatment as proxies for influence within social networks. Depending on their accuracy as proxies for influence, they may be able to indicate influence at the group level, which could be more predictive of the outcome of the diffusion process. An alternative type of intervention to simple seed fairness would thus be to incorporate network measures of influence into our definitions of fairness.

Consequently, we investigate the efficacy of interventions adhering to fairness definitions that incorporate network measures of influence. Intuitively, these interventions seed each group such that they possess equal amounts of influence within the network (assuming the network measure is a good proxy for influence). Our analysis probes the extent to which these interventions encapsulate information about the system in which diffusion takes place and how consistently they produce equitable outcomes.

In working toward this analysis, we first need to define new influence-based network interventions, necessitating group-level aggregations of our individual-level network influence measures.

Group-Level Network Measures: Integrating group-level influence into a framework for group fairness definitions necessitates computing network influence metrics at the group-level. To generate such group-level network metrics from the previously defined individual-level network measures (subsection 2.1), we must first select a cohort of individuals — either encompassing the complete group population or a specific subset of the group. We also must aggregate the individual-level network metrics to generate a consolidated measure, indicative of each group's overarching influence in the network based on a particular subset.[11]

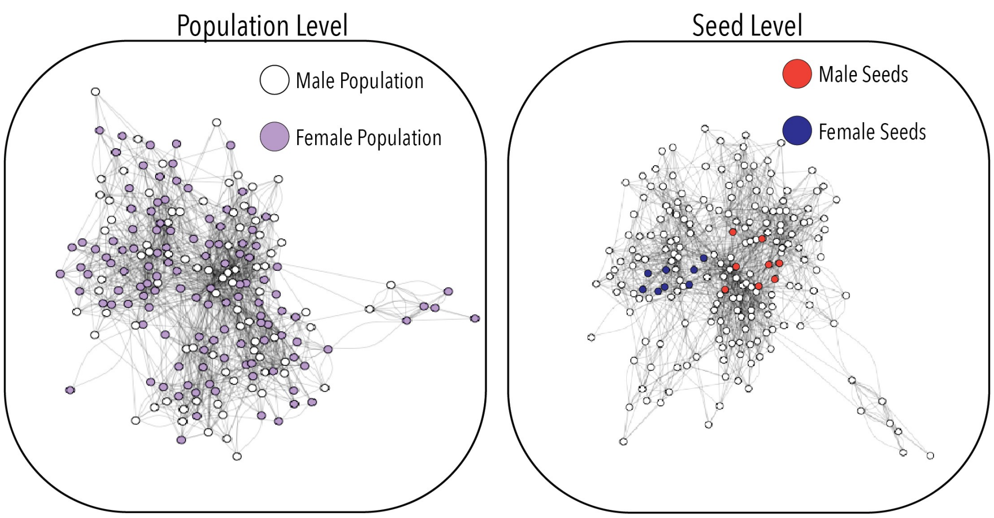

Figure 3: Population-Level Measures vs Seed-Level Measures

Selecting an exhaustive set of subsets of individuals for group measures poses challenges. Within the context of information diffusion, our primary concern revolves around the designation of the seed set, as well as the consideration of the entire population, in relation to the diffusion process's outcome. Consequently, we focus on two distinct subsets within each group: the seed set and the group's population. Representing minima and

[11] It is important to note that our focus is not on the quantity of connections one group extends to the other without redundancy. Hence, prevailing network measures tailored for inter-group dynamics remain beyond our current purview.

maxima subset sizes, the seed set and the population set function as reference benchmarks for diffusion trajectory. For the seed set, individual-level network metrics are assessed for seeds within each group and subsequently aggregated to derive a seed-set measure of influence for each group — characterizing the aggregate influence of each group's seed set. Conversely, at the population level, individual network metrics encompassing the complete group population are aggregated, generating a population-level network measure of group influence — or the collective influence inherent to the group's entire population.

Figure 3 illustrates the extent of both population-level and seed-level measures. The figure on the left depicts a network where individual-level network metrics for every node within the population are considered for both male and female villagers. This visualization represents an actual village (village 1), with purple nodes indicating female villagers and white nodes indicating male villagers. For population-level metrics, every individual's network measure of influence is integrated into the group's aggregate metric. Conversely, the figure on the right emphasizes only nodes designated as seeds. For clarity, both male and female villagers are represented by eight seeds each (e.g., village one has an actual seed set of size 16), where red nodes indicate male villagers and blue nodes indicate female villagers. At the seed level, the metrics of only the seeds are considered and subsequently aggregated. It is important that for both population and seed-level network metrics, equality is not merely a function of the quantity of seeds allocated to each group. Instead, equality is governed by the aggregated individual network metrics belonging to each group.

In this study, we consider two simple approaches to aggregation: 1.) computing the sum of individuallevel measures, and 2.) computing the average of the measures. When computing the sum, the aggregation preserves the overall magnitude or total quantity of the measures within each subset (either the seed set or population set). Conversely, the average represents the typical or central value within each group.

Total Influence: Computing group-level aggregates of the individual-level network measures is most straightforward for seed-level network metrics. For a given individual-level network metric of influence, denoted as x, the total influence inherent to a group's seed-set is computed as the summation of individuallevel network metrics, represented formally as:

$$\sum_{i=1}^{K} x_i,$$

where x_i denotes the individual-level network metric for individual i and K denotes the total count of individuals within the seed set. To define *total seed influence* distinctly for male and female seed sets, let k_M and k_F denote the seed sets for male and female villagers, respectively. The metrics are defined as:

$$\text{Male Seed Total Influence} = \sum_{i=1}^{k_M} x_i \tag{1}$$

$$\text{Female Seed Total Influence} = \sum_{i=1}^{k_F} x_i \tag{2}$$

By distinguishing these measures, we can examine the aggregated influence contributions from each gender-specific seed set. As an example, let us consider degree as our chosen network measure of influence for village one. Upon allocating an equal number of seeds to both male and female villagers the total influence attributable to the male seed set (with male villagers representing 8 out of the 16 possible seeds) amounts to 330. Conversely, the total influence for the female seed set, stemming from their 8 seeds, is 185. Evaluating the equality of these group-level measures will be detailed in the forthcoming section. However, we highlight an important observation at the foundation of these group-level network measures: even with equality in the number of seeds allocated to each group, we see a disparity in the aggregate influence within their respective seed sets. This, of course, is dependent on the assumption that degree serves as a strong indicator of influence (which it might not).

The total influence at the population level for each group, like the seed-set, requires computing the sum of individual-level network metrics across all members of each group. Let n_M and n_F denote the size of the sets of individuals for the male and female villager populations, respectively. The *total population influence* can then be defined as:

$$\text{Male Pop Total Influence} = \sum_{i=1}^{n_M} x_i \tag{3}$$

$$\text{Female Pop Total Influence} = \sum_{i=1}^{n_F} x_i \tag{4}$$

By defining these equations, we can quantify the influence distribution across the entire male and female populations within the network. For illustration, let us once again use degree as our selected network influence metric for village one. The computed total influence for the male villager population is 1,331, whereas for the female villager population, it is 1,451. In the next section, we will provide details on how to assess equality in relation to seed selection. One important observation to note is that the total influence metric for the female villager population at the population level exceeds that of the male villager population. This disparity can be attributed to the composition of village one, wherein the female villager population is larger than the male villager population.

Average Amount of Influence: A complementary method of aggregation is to compute the mean of individual-level network metrics, yielding group-level metrics aimed at the central tendencies of individuallevel metrics in the group. Specifically, these become the *average seed set influence* and the *average population set influence*. For a given individual-level network metric, denoted as x, the average influence attributed to a specific set corresponds to the expected value of x for that set. This can be formally defined as:

$$\frac{1}{K}\sum_{i=1}^{K} x_i ,$$

where x_i represents the individual's network metric, and K is the size of the set. Analogous to the total influence metrics, the mean influence for seed and population sets can be defined as:

$$\text{Male Seed Average Influence} = \frac{1}{k_M}\sum_{i=1}^{k_M} x_i \quad (5)$$

$$\text{Female Seed Average Influence} = \frac{1}{k_F}\sum_{i=1}^{k_F} x_i \quad (6)$$

Like before, to illustrate these measures, let us use degree as our selected network influence metric for village one. If we allocate an equal number of seeds to both male and female villagers (8 and 8, respectively), the male average seed influence is 41.2 and the female average seed influence is 23.1.

Similarly, we can formulate a group network measure for population using the mean as our aggregation approach, defined as follows:

$$\text{Male Pop Average Influence} = \frac{1}{n_M}\sum_{i=1}^{n_M} x_i \quad (7)$$

$$\text{Female Pop Average Influence} = \frac{1}{n_F}\sum_{i=1}^{n_F} x_i \quad (8)$$

In this definition, n_M and n_F continue to represent the total count of male and female villagers within the population, respectively. Using degree as our chosen metric of influence for village one, we find that the mean influence for the male villagers' population is 16, whereas for the female villagers, it is 12. These figures represent the average degree for each population group.

In the next section, we provide details on how to evaluate the equality of these group-level network measures to produce measures of group fairness for network interventions.

Defining and Evaluating Fair Seed Sets for Network Interventions: In the previous section, we stipulated that fair network interventions would ensure that each group holds equivalent influence based on a particular group-level aggregation of an individual-level network metric. Our interpretation of fairness changes depending on whether we select a seed-level or a population-level measure, and whether we aggregate individual-level network measures into a total value or an expected value. Consequently, we introduce four general fairness definitions for each group-level aggregation metrics, allowing any individual-level network metric to be incorporated as the influence measure.

The fairness definitions are specific to facilitated diffusion processes and presuppose that an intervention entails the allocation and distribution of *k* seeds across distinct groups. For our group-level network measures, which assess either the total or average influence of the seed set, the conversion into operational fairness definitions is fairly straightforward. Across all seeding allocations, individuals are selected from each group beginning with the villager exhibiting the maximal individual-level value for the chosen influence metric, proceeding thereafter in descending order.

Seed-Level Fairness Definitions: We define *seed total fairness* as an intervention wherein each group maintains an equivalent total influence within their respective seed sets. Formally, this implies that the sum of individual-level influence metrics within the male villager seed set (k_M) should be equal to the sum within the female villager seed set (k_F), as measured by Equations (1) and (2). This can be formally expressed as:

$$\sum_{i=1}^{k_M} x_i = \sum_{i=1}^{k_F} x_i, \qquad \text{(Seed Total Fairness)}$$

where x_i represents an individual-level network influence measure.

In a similar vein, *seed average fairness* defines an intervention where each group's seed set possesses equal mean influence. This is measured through Equations (5) and (6) and defined formally as:

$$\frac{1}{k_M}\sum_{i=1}^{k_M} x_i = \frac{1}{k_F}\sum_{i=1}^{k_F} x_i. \qquad \text{(Seed Average Fairness)}$$

It is important to note that *seed average fairness* does not account for the cardinality of seed allocations to each respective group; rather, the definition solely constrains the central tendency of the group-level network measure within each group's seed set. This aspect will be considered more closely in the analysis. Conversely, *seed total fairness* implicitly takes into account the quantity of seeds allocated to each group, given that the group-level metric is additive.

Population-Level Fairness Definitions: For population-level network measures, the fairness definitions require distinct specifications. Within the context of our seeding fairness paradigms, we aim to constrain seed selection according to a specific fairness criterion using a network metric that accounts for the influence in each group's entire population. We therefore interpret the population-level network measure as a weighting factor applied to the quantity of seeds allocated to each group. Formally, let s_M and s_F represent the count of male and female villagers selected as seeds, respectively, and let Ω_M and Ω_F denote the population-level network metrics corresponding to male and female villager populations. Our fairness definition necessitates the selection of s_M and s_F such that,

$$\Omega_M \cdot s_M = \Omega_F \cdot s_F \tag{9}$$

It is worth noting that this equation is equivalent to $\frac{\Omega_M}{\mathbf{\Omega}} \cdot s_M = \frac{\Omega_F}{\mathbf{\Omega}} \cdot s_F$, where $\Omega_M, \Omega_F \in \mathbf{\Omega}$ and $\mathbf{\Omega}$ signifies the aggregate population-level network metric for the entire village (including both male and female villager groups). Conceived this way, the population-level fairness definition can be interpreted as equating the intrinsic influence between groups. This definition mandates that the influence conferred upon each group via seed distribution be proportional to a fixed coefficient, in this case, the population-level network measure, thereby counteracting any inflation or dilution of inherent group-level influence as a function of seed allocation.

The general fairness definition outlined above can be translated into distinct fairness criteria for seed selection, based on both total and average population-level network metrics. The formal expression for *pop total fairness* is as follows:

$$\Omega_M \cdot s_M = \Omega_F \cdot s_F, \text{ where } \Omega_M = \sum_{i=1}^{n_M} x_i \text{ and } \Omega_F = \sum_{i=1}^{n_F} x_i. \tag{Pop Total Fairness}$$

Here, x_i denotes the individual-level network measure attributable to the i^{th} individual. This fairness definition requires that each seed acts as a representative of their group's cumulative influence. Consequently, each group should be proportionately represented in the seed set, weighted by the total influence intrinsic to that group.

Pop average fairness is formulated in a similar manner:

$$\Omega_M \cdot s_M = \Omega_F \cdot s_F, \text{ where } \Omega_M = \frac{1}{n_M} \sum_{i=1}^{n_M} x_i \text{ and } \Omega_F = \frac{1}{n_F} \sum_{i=1}^{n_F} x_i. \tag{Pop Average Fairness}$$

Again, x_i represents the individual-level network measure associated with the i^{th} individual. This fairness criterion can be interpreted analogously to its total counterpart: each seed functions as a representative of their respective group's mean influence. Therefore, fairness requires equitable representation within the seed set, as determined by the average influence of each group.

Selecting Seeds and Evaluating Seeding Fairness: To apply these fairness definitions as constraints within the framework of network interventions, we integrate them into a straightforward set of guidelines for seed selection:

1. Select a network measure for influence.
2. Choose one of the fairness definitions.
3. Select seeds for each group that maximize the selected fairness criterion.

Assessing the compliance of network interventions with these fairness definitions entails quantifying each group's contributions to the fairness definition. Due to their formulation as group-level influence measures, these fairness definitions are readily translatable fairness metrics. To illustrate, consider the definition of *seed average fairness*:

$$\frac{1}{k_M}\sum_{i=1}^{k_M} x_i = \frac{1}{k_F}\sum_{i=1}^{k_F} x_i .$$

We shall refer to each side of the expression of our fairness definition as the *seeding strategy metric* for male villagers and female villagers, respectively. That is, here, $\frac{1}{k_M}\sum_{i=1}^{k_M} x_i$ is the *seeding strategy metric* for male villagers under the *seed average fairness* definition. We can now compute fairness metrics quite easily by measuring the similarity of each group's *seeding strategy metric* under a specified fairness definition. To do so, we use a similarity metric implemented earlier to measure the fairness of diffusion outcomes between male and female villagers.

The statistical rate, *SR*, is defined as:

$$SR = \frac{\min(r)}{\max(r)}, \quad r \in \{X_M, X_F\},$$

where X_M and X_F signify values corresponding to the male and female villager populations, respectively. The computation of the statistical rate entails the division of the minimum measure by the maximum, thereby resulting in a similarity coefficient that ranges from 0 to 1, with 1 indicating perfect parity. The coefficient can be interpreted as representing the fraction of the larger measure that the smaller one constitutes, shedding light on the relative scale of the disadvantaged group in comparison to the privileged group. This metric is pertinent in the present context as we aim to compare the relative allocations of a given resource between distinct groups (2022).

To gauge the similarity of each group's *seeding strategy metric* (γ) for a specified fairness definition, we can employ the statistical rate. Let γ_M denote the *seeding strategy metric* for male villagers and γ_F for female villagers. As an example, for *seed average fairness*, we have:

$$\gamma_M = \frac{1}{k_M} \sum_{i=1}^{k_M} x_i \quad ,$$

$$\gamma_F = \frac{1}{k_F} \sum_{i=1}^{k_F} x_i \quad .$$

The similarity of the *seeding strategy metrics* is computed as:

$$Seed_{SR} = \frac{\min{(r)}}{\max(r)}, \quad r \in \{\gamma_M, \gamma_F\} \quad . \tag{Seed Stat Rate}$$

This generalized similarity metric is applicable to *seeding strategy metrics* for any of our specified fairness definitions. We will refer to this seeding fairness metric as the *seed stat rate*. Leveraging this general fairness metric, we can proceed with a straightforward approach to maximizing the fairness of a network intervention for a chosen fairness definition (guideline 3, above):

1. Initialize the available seed quota, k, for a village, dividing the total number of seeds equitably between male and female villagers, as k_M and k_F. When the total seed count is odd, allocate seeds to achieve the closest possible balance.

2. Begin seed selection by identifying individuals with the highest values based on the individual-level network measure used to approximate influence and proceed selection in descending order until k_M and k_F seeds are allocated to each respective group.

3. Use the *seed stat rate* metric to quantify the fairness of the initial seed allocation for the chosen fairness definition.

4. In instances where the male villager *seeding strategy metric* (γ_M) exceeds that of the female villagers', de-allocate the last male seed added to k_M and add the next highest-ranking female individual to k_F — based on the chosen individual-level network measure — who is not already included in k_F.

5. Conversely, if the female villager *seeding strategy metric* (γ_F) is greater, remove the most recently added female seed from k_F and incorporate the next highest-ranking male individual not already included in k_M.

6. Re-evaluate the *seed stat rate* to assess the adjusted seed allocation for fairness.

7. Iteratively execute steps 4-6 until the *seed stat rate* approximates 1, thereby indicating maximized fairness in seed allocation.

This procedure iteratively adjusts the seed allocation to achieve a state closest to ideal fairness, as quantified by the *seed stat rate* metric.

Evaluating Outcome Fairness: While the outcome fairness of a diffusion process is not within our direct control, it can be indirectly influenced through strategic network interventions. To examine the efficacy of these interventions in generating fair diffusion outcomes, a metric for evaluating outcome fairness is essential. As mentioned in subsection 2.5, we define outcome fairness as the condition where both male and female villagers have an equivalent probability of being informed. Formally, this is expressed as:

$$P(i \in O : i \in M) = P(j \in O : j \in F), \qquad \text{(Outcome Fairness)}$$

Here, i and j are individuals from the sets of male (M) and female (F) villagers, respectively, and O designates the set of villagers who have been informed.

To evaluate outcome fairness, we first compute the diffusion outcomes ω_M and ω_F, denoting the proportions of informed individuals within the male and female villager populations, respectively. We subsequently evaluate fairness by employing statistical rate to measure the similarity between these two outcome ratios:

$$Outcome_{SR} = \frac{\min(r)}{\max(r)}, \quad r \in \{\omega_M, \omega_F\}. \qquad \text{(Outcome Stat Rate)}$$

Using the *outcome stat rate* metric, we can evaluate the disparities in the outcomes for each group, thereby providing an assessment of the fairness achieved through our various network interventions.

3 Analysis of Fairness in Diffusion

In this study, we examine the effectiveness of a set of fair seeding interventions in achieving equitable outcomes. We focus on the reliability of thirteen distinct network influence metrics as candidates for fair seeding interventions. Each metric is tested under four distinct group-level fairness constraints, using set of models that emulate real diffusion processes observed in Uganda (Ferrali et al., 2020). Each model approximates the information diffusion process of one of fourteen villages where information about a PCT is spread among male and female villagers, beginning with an initial set of seeds (subsection 2.3). Though our models encapsulate features of the actual diffusion dynamics in these villages, the diffusion model's framework (subsubsection 2.3.1) is relatively straightforward. This suggests that our findings may have broader applicability across various scenarios, yet they may also be somewhat more optimistic with respect to achieving equitable outcomes compared to a real diffusion process where additional complexities and individual nuances come into play.

For each village, we apply a simulation for all thirteen of our network influence measures (subsection 2.1), utilizing a seeding approach dictated by one of our four fairness definitions: *seed total fairness*, *seed average fairness*, *pop total fairness*, and *pop average fairness* (subsection 2.5). In each simulation, we ensure that the size of the seed set is equal to the size used in the actual diffusion process of each village.

In assessing the effectiveness of fair seeding interventions in relation to outcome fairness, we adhere to a strict set of criteria (subsection 2.5). Our objective is to analyze the empirical distribution of seeding strategies and their capacity to yield outcome fairness. Consequently, the primary thrust of the analysis centers on success rates contingent upon specific intervention selection decisions. We look to understand the likelihood associated with potential decisions. To this end, we establish criteria detailing the probability of an outcome based on our capacity to opt for particular interventions:

1.) Given a random intervention selection, the probability associated with a fair outcome is determined by the ratio of fair outcomes to the size of the set of possible interventions; 2.) Opting for a maximized efficiency intervention yields a certain probability of manifesting a fair outcome; 3.) For specific conditions — such as village, network measure, or fairness definition — we can select the intervention that maximizes seeding fairness. We refer to these as "maximized fair interventions"

Additionally, hybrid decisions can be employed, combining elements of random selection with either maximized efficiency or maximized fair interventions, like randomly choosing a maximized fair intervention from among all the network measures of influence.

The purpose of clarifying our intervention-selection framework is to lend some analytical rigor to the analysis. We aim to investigate the probability of attaining equitable outcomes based on intervention choices. By comparing the likelihood of achieving a fair outcome through a fair intervention against the probabilities associated with maximized efficiency interventions and random choices, we gain a clearer perspective on the efficacy of fair interventions.

Our primary focus within the domain of fair interventions is directed towards those that are maximized fair interventions (e.g., *seed stat rate* closest to 1). We have two reasons for this rationale: 1.) We want to replicate a real decision-making environment wherein maximizing intervention fairness is assumed to be the best strategy for realizing equitable outcomes; and 2.) We have no prior information suggesting that fairer interventions might yield less equitable outcomes.

Although the focus on maximized fair, random, or maximized efficiency interventions may appear somewhat discretionary, it provides a practical lens to interpret the performance of the interventions in the diffusion process simulations. Absent this guideline, our assessment might inadvertently emphasize favorable outcomes derived retrospectively, potentially overstating the probability of achieving equitable outcomes via a fair intervention. For instance, within a specific village, and given a particular network measure and fairness definition, we could observe a highly equitable result from a very fair seeding intervention (e.g., statistical rates close to 1 for both criteria). However, if this seeding intervention is not the fairest intervention, we lack a good rationale to opt for a comparatively less fair intervention without prior knowledge of its potential to generate a more equitable outcome.

Nevertheless, we evaluate every possible allocation across male and female villagers for each village, network metric, and fairness definition. This facilitates an examination of three aspects of fair diffusion for our analysis: 1.) We can observe the relationship between predictability and the network measure, fairness definition, and case-specific attributes (e.g., village); 2.) We can check the sensitivity of small allocation variations on outcome fairness; and 3.) We can observe predictable relationships between outcome fairness and our allocations that are possibly overlooked by our current set of fairness definitions.

For every combination of village, network measure, fairness definition, and allocation pair, we conduct the diffusion process simulation 1,000 times. We then average these results to observe the more stable central tendencies associated with each village, network measure, fairness definition, and allocation pair. Additionally, as previously illustrated, we generate fairness results and cumulative rates for maximizing interventions pertaining to each network measure and village. This aids in structuring comparisons between the maximizing and fair interventions within our study, and provides benchmarks for defining improvements in outcome fairness.

The analysis is structured into four sections: 1.) A summary of fair intervention results; 2.) Efficacy of the maximized fair interventions; 3.) Discrepancies, sensitivities, and deviations of maximized fair interventions in relation to outcome fairness; and 4.) Potential implications of fair interventions on the collective efficiency of information spread.

The first section of the analysis offers summary statistics for all simulations. We observe the prevalence of equitable outcomes across all interventions, those specifically labeled as 'fair,' and those prioritizing maximum efficiency. Here, we observe, from a broad perspective, the frequency with which each network metric facilitated fair seeding and the subsequent realization of these fair interventions into equitable outcomes.

The second section targets the efficacy of the maximized fair interventions for each village, network metric, and fairness definition. We observe the frequency with which maximized fair interventions produce more equitable outcomes than their max efficiency counterparts, using the latter as a benchmark for improving outcome fairness. We also observe the success rate in producing equitable outcomes (e.g., *outcome stat rate* ≥ 0.9) from maximized fair interventions, and the robustness of various network metrics in generating equitable outcomes across multiple villages.

The third section of the analysis scrutinizes several nuances in generating outcome fairness: specifically, how sensitive outcome fairness is to minute changes in allocations. We also examine allocation and seeding fairness patterns that appear indicative of outcome fairness, that might not be captured by the current fairness definitions. In light of these observations, a fifth fairness definition and intervention strategy is conceptualized and implemented to correct a potential bias that occurs when *seed average fairness* is used under certain conditions.

Concluding the analysis, the final section investigates the potential costs of fair interventions and outcomes on the overarching efficiency of the diffusion process, relative to efficiency maximizing strategies. This has important implications, as choosing collective fairness might come at the cost of disseminating information to a broader segment of the populace, and vice versa. We examine costs associated with fair interventions that either do or do not yield fair outcomes, and fair outcomes that aren't necessarily a byproduct of fair interventions.

Summary of Fair Intervention Results: Table 5 enumerates frequencies of specific events across all allocations. The "Interventions" column delineates classes of interventions. It initiates with general categories, transitioning to frequencies associated with our network measures under "Main Strategies." In the context of these general categories, the table enumerates the aggregate count of interventions under "Total", as well as instances of fair outcomes where *outcome stat rate* ≥ 0.9, as indicated under "Outcomes." The succeeding column, "% Fair Outcomes," translates the raw count data into a percentage representation of events yielding a fair outcome.

In the "Main Strategies" section, the table details interventions achieving a *seed stat rate* ≥ 0.9 ("Fair Seedings") and the proportional representation of seedings deemed fair under this criterion, ("% Fair Seeding"). "Outcomes" are adjusted under the "Main Strategies" to account for the number of fair outcomes from among the count of fair seedings. "% Fair Outcomes" under "Main Strategies" translates the "Outcomes" counts into a percentage of the number of fair seedings. Distinct parameter choices for complex centrality (t = {2,4,6}) and influence maximization (p = {0.01,0.10,0.25,0.50}) are given individual rows in the table for more granularity. The tabulated data encompasses all allocation pairs, spanning all 14 villages, and each of the four fairness definitions.

Interventions	Total			Outcomes (SR ≥ 0.9)	% Fair Outcomes
All Allocations/All Strategies	11,804			2,143	18.15%
Maximization Strategies	196			20	10.99%
Fair Seeding Strategies (SR ≥ 0.9)	1,590			200	12.58%
Main Strategies		Fair Seedings (SR ≥ 0.9)	% Fair Seeding		
Complex Centrality	2,724	510	18.72%	84	16.47%
t = 2	908	191	21.03%	40	20.94%
t = 4	908	201	22.14%	35	17.41%
t = 6	908	118	12.99%	9	7.62%
Influence Maximization	3,632	307	8.45%	34	11.07%
p = 0.01	908	75	8.26%	8	10.67%
p = 0.10	908	76	8.37%	8	10.53%
p = 0.25	908	76	8.37%	7	9.21%
p = 0.50	908	80	8.81%	11	13.75%
Degree Centrality	908	70	7.71%	4	5.71%
Eigenvector Centrality	908	56	6.17%	6	10.71%
Percolation	908	134	14.76%	9	6.72%
Closeness	908	189	20.81%	12	6.35%
K-Core	908	266	29.29%	36	13.53%
Betweenness	908	58	6.38%	15	25.86%

Table 5: Frequencies for all interventions

This section presents a summary and analysis of fairness simulations conducted across different network metrics. The aim of this analysis is to observe the efficacy of fair interventions with respect to producing equitable outcomes, and compare these likelihoods against choosing randomly or choosing a maximized efficiency strategy. A baseline comparison for *seed stat rate* and *outcome stat rate* is set at 0.9 to represent a fairness threshold that surpasses the best average *outcome stat rate* across villages associated with an efficiency maximizing strategy (Complex Centrality t = 2). A statistical rate of 0.9 signifies that the underprivileged group attained 90% of what the privileged group achieved, reflecting a high degree of equity within the system.

The aggregate numbers reveal that fair interventions generally exhibit a higher success rate for fair outcomes compared to maximizing strategies. Despite this observation, a significant number of fair outcomes stem from unfair interventions. The success rate of a fair outcome, with a randomly chosen allocation of seeds among male and female villagers and a random network measure, is 18.15%. This rate surprisingly underscores that a randomly selected allocation and network measure tends to yield more equitable outcomes compared to our fair interventions.

The table shows fair interventions overall to not be especially predictive of outcome fairness. Even if we include *simple seed fairness* (see Equation Simple Seed Fairness), which evaluates the proportional representation of each group, only 8.33% of the fair interventions (*seed stat rate* ≥ 0.9) achieve a fair outcome (*outcome stat rate* ≥ 0.9).

Observing specific main strategies, the network measures exhibit varying potentials for achieving fairness at the seeding stage. Notably, Betweenness, Eigenvector Centrality, Degree, and Influence Maximization generally exhibit lower rates of fair interventions, whereas Complex Centrality, Closeness, and K-Core demonstrate more capacity to achieve fair seed sets. These variations may stem from the inherent distributions of the network measures between male and female villagers.

No clear pattern emerges between the prevalence of attaining fair interventions and the likelihood of achieving equitable outcomes from these interventions. Degree Centrality, Eigenvector Centrality, Percolation, and Closeness all register lower rates of equitable outcomes from fair interventions compared to randomly choosing a Maximization Strategy. Contrarily, Complex Centrality, K-Core, and Betweenness portray higher rates, with only Complex Centrality ($t = 2$) at 20.94% and Betweenness at 25.86% exceeding the rates of achieving a fair outcome from a fair intervention than selecting an allocation and network measure at random.

The table highlights a number of baseline rates for achieving equitable outcomes. If we chose fairness definitions at random, we might only be able to achieve a fairer outcome than a completely random allocation and network measure in the long run with 2 of our 13 network influence measures. Outside of choosing randomly, we can choose the fairest intervention for each village, network measure and fairness definition. This will be the focus of the next section.

Efficacy of the Maximized Fair Interventions: In the previous section, we explored the success rate for attaining a fair outcome (where *outcome stat rate* is ≥ 0.9) from a fair intervention (randomly selected across all network metrics or within a specific network measure) compared to choosing an allocation and network measure entirely at random, or randomly selecting a maximizing strategy. In this part, we investigate frequencies associated with maximized fair interventions. These are the interventions for each network measure and village that achieved the highest *seed stat rate* and represent a subset of the fair interventions in Table 5. These are the interventions we would likely choose if our selection was not random, as selecting a fair seeding intervention that is not the fairest lacks a compelling reason, absent prior knowledge of its capacity to yield a more equitable outcome.

Overall, our concern extends beyond the ability of fair interventions to generate almost perfect fair outcomes. We are equally interested in the likelihood of a fair intervention improving outcome fairness compared to traditional maximizing strategies. Examining all combinations of village, network metric, and fairness definition, we have 728 scenarios in which we compute a maximized fair intervention. In 467 of these, or 64.15%, the maximized fair interventions yielded more equitable outcomes than their corresponding maximizing strategy. The average *outcome stat rate* for maximizing strategies across all villages and network measures was 0.649, whereas for maximized fair interventions, it was 0.695. Although this suggests that fair interventions should generally yield better outcome fairness than maximizing strategies, the improvement is modest. An average *outcome stat rate* of 0.695 implies that the underprivileged group (female villagers) attains only 69.5% of the information spread that male villagers achieve.

We also notice considerable variation across different network metrics regarding the enhancement of outcome fairness compared to maximizing strategies. For instance, in the case of Complex Centrality ($t = 2$), the average *outcome stat rate* for maximizing strategies across all 14 villages stands at a substantial 0.889. Given its proximity to a completely equitable outcome, it proves more challenging to improve outcome fairness for this measure compared to Degree or Eigenvector Centrality, which have the two lowest average *outcome stat rates* for maximizing strategies at 0.580 and 0.543, respectively.

Table 6 presents success rates for the maximized fair interventions. These interventions have achieved the highest *seed stat rate* values for each village, network measure, and fairness definition. For “All Strategies”, the table provides aggregate frequencies for the maximized fair interventions. Subsequent rows denote individual network measures and their corresponding performance frequencies. The “Fair Seeding” column denotes the count of interventions that attained a *seed stat rate* of 0.9 or above. The “% Fair Seeding” column translates this figure into a percentage of the fair seeding count. The ”Outcomes” column specifies the number of instances where an *outcome stat rate* of 0.9 or above was realized. The “% Fair Outcomes” column represents the percentage of the fair seeding interventions that also resulted in a fair outcomes Parenthesized values provide frequency and percentage adjustments, taking into account only those interventions that outperformed their corresponding maximized efficiency intervention in terms of outcome fairness.

We observe that the success rates for fair outcomes with maximized fair interventions are even lower than

Max Fair Interventions		Fair Seeding (SR ≥ 0.9)	% Fair Seeding	Outcomes (SR ≥ 0.9)	% Fair Outcomes	
All Strategies	728	565	77.61%	55 (39)*	9.73%	(6.90%)
Complex Centrality	168	143	85.12%	13 (11)	9.09%	(7.69%)
t = 2	56	47	83.93%	7 (6)	14.89%	(12.77%)
t = 4	56	48	85.71%	4 (3)	8.33%	(6.25%)
t = 6	56	48	85.71%	2	4.17%	
Influence Maximization	224	168	75.00%	19 (9)	11.31%	(5.36%)
p = 0.01	56	40	71.43%	4 (2)	10.00%	(5.00%)
p = 0.10	56	42	75.00%	5 (3)	11.90%	(7.14%)
p = 0.25	56	43	76.79%	5 (3)	11.63%	(6.98%)
p = 0.50	56	43	76.79%	5 (1)	11.63%	(2.33%)
Degree Centrality	56	40	71.43%	1	2.50%	
Eigenvector Centrality	56	38	67.86%	5	13.16%	
Percolation	56	46	82.14%	2 (0)	4.35%	(0.00%)
Closeness	56	45	80.36%	2 (0)	4.44%	(0.00%)
K-Core	56	46	82.14%	5	10.87%	
Betweenness	56	39	69.64%	8	20.51%	

Table 6: Frequencies for maximized fair interventions

those observed in Table 5. Across all network metrics, only 39 maximized fair interventions attain *seed stat rates* ≥ 0.9, *outcome stat rate*s ≥ 0.9, and yield more equitable outcomes compared to corresponding maximizing strategies. Including fair interventions that did not surpass their maximizing strategy counterparts in equitable outcomes raises the count of fair interventions to 55. Moreover, if the *seed stat rate* being ≥ 0.9 is not a concern, focusing only on achieving *outcome stat rate*s ≥ 0.9 increases the count of maximized fair interventions that generate equitable outcomes to 62.

These frequencies continue to reflect quite low success rates for equitable outcomes. Selecting a maximized fair intervention at random presents a 9.73% chance of attaining an equitable outcome (only 6.9% if we necessitate improving outcome fairness over maximizing strategies). It is observed that among the set of maximized fair interventions, only Betweenness surpasses the likelihood of attaining a fair outcome compared to randomly choosing an allocation and network measure, which carries a success rate of 18.15% in achieving a fair outcome (Table 5).

Filtering the maximized fair interventions to those that achieve *seed stat rates* ≥ 0.9, *outcome stat rate*s ≥ 0.9, and yield more equitable outcomes than corresponding maximizing strategies, we notice a downward adjustment for Complex Centrality and Influence Maximization. This suggests that these network measures more easily facilitated equitable outcomes for certain villages, regardless of whether the intervention was fair or maximized when seeding. Indeed, Percolation and Closeness offer no beneficial maximized fair interventions as the limited ones (2) that attain *seed stat rate* ≥ 0.9 and *outcome stat rate* ≥ 0.9 do not foster fairer outcomes than their corresponding maximizing strategies.

In relation to this, we notice varying ease across villages in attaining fair outcomes. Here, Betweenness emerges as the most robust strategy, yielding fair outcomes from maximized fair interventions in 6 out of the 14 villages, accounting for 42.28% of the villages, all under *seed total fairness*. Complex Centrality (t = 2) stands as the second most reliable strategy, generating fair outcomes in 4 out of the 14 villages, or in 28.57% of the network settings all under *seed average fairness*. Given that for both of these network measures, maximized fair interventions attain fair outcomes under singular Fairness Definitions, this effectively alters the anticipated probability of their success to 42.28% and 28.57% for Betweenness and Complex Centrality (t = 2) respectively, if they were selected as interventions across all villages. This improvement is notable compared to an 18.15% rate of success for attaining a fair outcome by randomly choosing an intervention.

Examining Complex Centrality in a broader context, it yields fair outcomes in 10 out of the 14 villages, equivalent to 71.42%. In aggregate, maximized fair interventions managed to produce an equitable outcome (*outcome stat rate* ≥ 0.9) in 11 out of the 14 villages, totaling 78.57%. Nevertheless, Villages 5, 7, and 9 were restricted to the highest average *outcome stat rate*s of 0.720, 0.487, and 0.799, respectively. In this scenario, it is evident that certain networks pose significant barriers to achieving fair outcomes. The likelihood of attaining a fair outcome from a maximized fair intervention in these villages effectively dwindles to zero. Hence, the probability of attaining fair outcomes in these villages with a randomly selected intervention is vastly higher — given that every village possesses at least one allocation with *outcome stat rate*s meeting or exceeding 0.9.

Choosing maximized fair interventions does not markedly enhance the chances of realizing a fair outcome. However, they do enhance fairness on average compared to maximizing strategies, which holds significance. The most effective maximized fair intervention is Betweenness and Complex Centrality, which effectively each have success rates of 42.28% under *seed total fairness* and 28.57% under *seed average fairness*, respectively, if applied across all villages. One notable distinction is that Complex Centrality does not readily attain fairer outcomes compared to its performance as a maximizing strategy. This could imply that its ability to yield fair outcomes is substantially affected by its inherent characteristics within the networks. This situation is not as transparent in the case of Betweenness, which typically shows a highly uneven distribution between male and female villagers and does not produce equitable outcomes as a efficiency maximizing strategy (average *outcome stat rate* = 0.602).

In the next section, we delve into potential issues related to our fairness definitions that curtail their capacity to predict equitable outcomes effectively through maximized fair interventions.

Discrepancies, Sensitivities, and Uncertainty of Maximized Fair Interventions: Each of the 14 villages, when paired with one of the 13 network metrics and one of the four fairness definitions, yields a total of 728 possible cases. Among these, a certain allocation of seeds led to an *outcome stat rate* of ≥ 0.9 in 544 cases, amounting to 74.72%. As observed in the previous section, a mere 6.9% of the maximized fair interventions with a *seed stat rate* of ≥ 0.9 also resulted in an *outcome stat rate* of ≥ 0.9 and exhibited fairer outcomes than the corresponding maximized efficiency strategy for the specific network measure. For every village, there exited at least one allocation for an intervention attaining an *outcome stat rate* of ≥ 0.9. Although village eight had only one intervention (with one network metric) that reached or exceeded an *outcome stat rate* of 0.9, in 8 villages or 57.14%, an allocation existed where the *outcome stat rate* was ≥ 0.9 for all four fairness definitions and all 13 network metrics.

This highlights the apparent weaknesses in our fair interventions, given their limited predictive capacity for equitable outcomes despite the existence of potential allocations that could foster such results. This issue seems partly linked to the practical constraint that we must maximize the fairness of our intervention, or else pick randomly or use a maximized efficiency strategy. The *outcome stat rate* demonstrates sensitivity to minor allocation differences. We find that across our 13 network measures, a solitary change in seed allocation can cause an average oscillation in the *outcome stat rate* between 0.0491 and 0.0958. This susceptibility is even more pronounced across villages, with the *outcome stat rate*, on an average, changing between 0.01379 and 0.2451 with a single alteration in seed allocation. For all 728 scenarios, the average corresponding *outcome stat rate* arising from a maximized fair intervention stands at 0.695, while the average best *outcome stat rate* for each scenario is 0.913.

Upon maximizing fairness for our interventions under each fairness definition, the results consistently generated divergent allocation pairs compared to the optimal (for equitable outcomes). Under *pop total fairness*, our maximized fair interventions typically identified an allocation deviating by an average of 5.96 seeds from the most equitable outcome allocation. This scenario translates to one group, on average, having nearly six seeds in excess, while the other group has six seeds fewer than they should. Similarly, under *pop average fairness*, maximizing the fairness of the intervention on average leads to an allocation disparity of 4.83 seeds from the optimal. For *seed total fairness*, the average deviation is 4.5 seeds. Notably, *seed average fairness* shows a significantly higher average discrepancy of 9.11 seeds from the optimal allocation.

Across the 728 scenarios, the maximized fair intervention on average gives 8.83 seeds to male villagers and 8.38 seeds to female villagers, when the optimal intervention for equitable outcomes on average gives 2.81 seeds to male villagers and 14.41 seeds to female villagers. These values averaged across all fairness definitions are significantly affected by *seed average fairness* for which we on average give 11.63 seeds to male villagers and 5.89 seeds to female villagers.

The *seeding strategy metrics* for *seed total fairness*, *pop total fairness*, and *pop average fairness* of each network measure underestimate the number of seeds that should be allocated to female villagers, which suggest that certain characteristics of the male and female villager populations with respect to passing information are not captured by each group's distribution of the network measure of influence. Perhaps important network characteristics or aspects of how the characteristics are distributed between groups are overlooked by using the total or average of the individual-level network measures, or perhaps there are non-network factors such as the likelihood of adoption leading to such significant underestimates. Such characteristics represent challenges for designing fair interventions.

Although seeding strategy metrics for *seed total fairness*, *pop total fairness*, and *pop average fairness* may yield underestimations for fair interventions, they all allocate more seeds to female than male villagers, aligning with the average optimal allocation. In contrast, *seed average fairness* allocates more seeds to male villagers, imposing seed fairness that diverges from the optimal allocation average. This discrepancy highlights a potential issue in the way *seed average fairness* interprets the proper allocation, emphasizing the need for reassessment and adjustment to more closely align with optimal allocation strategies.

Table 7 details the frequencies with which maximized fair interventions attained outcomes that were either more or less fair than those resulting from their corresponding maximized efficiency interventions. The first row captures frequencies for all such maximized fair interventions, while subsequent rows disaggregate each column according to specific fairness definitions. Percentages enclosed in parentheses within the total row represent the proportion of interventions that were either fairer or less fair when compared with their corresponding maximized efficiency interventions. Parenthesized percentages in the subsequent rows indicate the share of the column's total attributable to each respective fairness definition.

	Fairer Outcomes	Less Fair Outcomes
Total	467 (64.14%)	261 (35.85%)
Seed Average	73 (15.63%)	109 (41.76%)
Seed Total	133 (28.48%)	49 (18.77%)
Pop Average	138 (29.55%)	44 (16.86%)
Pop Total	123 (26.33%)	59 (22.61%)

Table 7: Frequency of iterations by fairness definition

Maximized fair interventions yield more equitable results than maximized efficiency interventions in 64.14% of the 728 scenarios. *Seed total fairness*, *pop total fairness*, and *pop average fairness* show relatively consistent outcomes, producing comparable numbers of fairer and less fair results. This consistency implies that these strategies may not be fundamentally flawed, but rather they are not particularly predictive for ensuring equitable outcomes. On the other hand, *seed average fairness* significantly underperforms, marking the lowest number and percentage of fairer outcomes and the highest number and percentage of less fair outcomes.

These findings are perplexing. In the prior section we explored a subset of 39 scenarios in which the maximized fair intervention *seed stat rate* ≥ 0.9, resulting in an *outcome stat rate* ≥ 0.9 and also yielding more equitable outcomes than the corresponding maximized efficiency strategies. Interestingly, the largest portion (18/39) were maximized *seed average fairness* interventions, with 11/39 being *seed total fairness*, 8/39 being *pop average fairness*, and a mere 2/39 being *pop total fairness*.

Figure 4 illustrates the proportion of interventions associated with each fairness definition as a function of an R^2 measure relating *seed stat rate* and *outcome stat rate*. With an increase in R^2, we observe an increasing predominance of interventions aligning with the *seed average fairness* definition. This trend suggests that many interventions under the *seed average fairness* definition exhibit a consistent relationship between *seed stat rate* and *outcome stat rate*. Consequently, it is possible that these interventions demonstrate a predictable relationship with equitable outcomes, yet are overlooked in selection because their *seed stat rates* do not positively associate. In many of these cases, we see the prevalence of a pattern where as the *seed stat rate* diminishes, the *outcome stat rate* correspondingly improves.

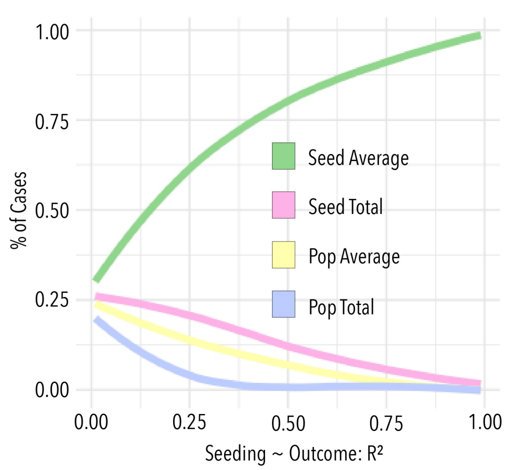

Figure 4: R-squared in relation to fairness definition

The issue at hand stems from the limitations of the *seed average fairness* definition, which considers only the central tendency or expected value of a specific network measure within each group's seed set, without acknowledging the quantity of seeds allocated to each group. This method aims to balance the averages of each seed set, but it overlooks the impact of the distribution of the network measures.

For instance, if the distribution of a network measure among male villager seeds dominates that among the female villager seeds, the only way to lower the average network measure for the male villager seed set would be to allocate more seeds to male villagers. This situation arises because seeds are allocated based on the individual ranking with respect to the network measure, meaning any additional individual will have a lower network measure compared to those already in the seed set. Consequently, under such conditions, adding a seed decreases the group's average network measure, while removing one increases it. *Seed average fairness*, therefore, works as intended only when the distributions of the network measures across groups are relatively balanced.

A possible solution to this problem could be the creation of an alternative fairness definition. This definition would incorporate a weight to each group's seed set average, compensating for the number of seeds allocated to each group. This can be achieved by introducing a weight that effectively measures each group's aggregate coverage of the network measure within the population.

In the previously defined *seed average fairness* for a particular group:

$$\frac{1}{K}\sum_{i=1}^{K} x_i\,,$$

x_i denotes the individual's network metric, and K is the size of the set for a specific group.

To adjust for each seed set's size, this group-level measure is multiplied by a weight, ω, calculated as:

$$\omega = \frac{\sum_{i=1}^{K} x_i}{\sum_{i=1}^{N} x_i}$$

By this method, the weight ω effectively takes into account the proportion of each group's aggregate coverage of the network measure within the entire population, thus ensuring an allocation considers the average of the seed set with respect to how much aggregate network influence this seed average represents in the entire network. Here, $\sum_{i=1}^{K} x_i$ represents the total of individual network measures in the seed set K, while $\sum_{i=1}^{N} x_i$ denotes the total of individual network measures in the entire village population. Effectively, the seed average for a particular network measure is weighted by the proportion of the each group's total network measure in the population that their average seed value represents.

In this scenario, a small number of seeds set might exhibit a high average value for a particular network measure, as each seed is added based on individual rank concerning the network measure. However, a smaller seed set size will represent a lesser fraction of the total network measure present in the network, and therefore, a smaller weight will be assigned to the seed average of the network measure. This adjustment permits the *seed stat rate* to positively correlate with the *outcome stat rate* for many interventions that previously, under the *seed average fairness*, demonstrated a high R^2 in the relationship between *seed stat rate* and *outcome stat rate*.

We will refer to this new fairness definition as *seed average × coverage fairness*, defined as:

$$\frac{\omega_M}{k_M}\sum_{i=1}^{k_M} x_i = \frac{\omega_F}{k_F}\sum_{i=1}^{k_F} x_i \quad , \qquad \text{(Seed Average × Coverage Fairness)}$$

In this equation, ω_M and ω_F denote the coverage weights for the male and female seed sets, respectively. This fairness constraint is designed to equalize the average influence of each group's seed set, proportional to the fraction of total influence that each group's seed set holds within the entire village.

Table 8 presents the average *outcome stat rates* for all villages by comparing the statistical rates of maximized fair interventions of our initial *seed average fairness* interventions against the updated *seed average × coverage fairness* interventions. The "Old" column lists the average *outcome stat rates* associated with *seed average fairness*, while the "New" column presents *outcome stat rates* associated with *seed average × coverage fairness* for maximized fair interventions. The column labeled "Improvement/Loss" reflects the fractional change in the "New" *outcome stat rate* compared to the "Old" *outcome stat rate*.

	Old	New	Improvement/Loss
Complex Centrality			
t = 2	0.845	0.673	-0.172
t = 4	0.803	0.653	-0.150
t = 6	0.694	0.694	0.000
Influence Maximization			
p = 0.01	0.565	0.840	+0.274
p = 0.10	0.569	0.841	+0.271
p = 0.25	0.574	0.856	+0.283
p = 0.50	0.562	0.833	+0.271
Degree Centrality	0.542	0.843	+0.301
Eigenvector Centrality	0.523	0.841	+0.318
Percolation	0.719	0.722	+0.003
Closeness	0.544	0.707	+0.164
K-Core	0.791	0.685	-0.106
Betweenness	0.566	0.836	+0.270

Table 8: Change in *outcome stat rate* with Coverage Metric

We notice that most interventions for each network metric significantly improve under the *seed average × coverage fairness*. Strategies involving Influence Maximization, along with Degree, Eigenvector, and

Betweenness, all display growth in their average *outcome stat rate* values, with improvements ranging between 27.4% and 31.8%. However, a notable trade-off is observed. Strategies involving Complex Centrality and K-Core interventions yield lower *outcome stat rate* averages under this fairness definition. Recall that female villagers actually held more dominant distributions of Complex Centrality $t = 2$ and $t = 4$ (see subsection 2.2). This alternate fairness definition appears to bolster the *outcome stat rate* for network measures where male villagers had dominant distributions, highlighting the nuanced impacts of different fairness definitions on varied network metrics.

Seed average × *coverage fairness* exhibits a success rate of 59.57% in generating fairer outcomes as maximized fair interventions compared to corresponding efficiency maximizing strategies. This is notably higher than the 40.11% success rate observed under the original *seed average fairness* definition.

Table 9 displays the top 10 best maximized fair interventions, averaged across all villages. The first column lists the network measure, followed by the fairness definition in the second column. The third column shows the average *outcome stat rate* for the maximized fair interventions under the given fairness definition across all 14 villages. The fourth column provides the average *outcome stat rate* for each network measure as a maximized efficiency intervention. The final column shows the percentage improvement or loss in *outcome stat rate* over the maximized efficiency intervention of selecting the fair intervention.

	Fairness Definition	Avg. Outcome SR	Max Avg. SR	% Improvement/Loss
Influence Maximization (p = 0.25)	Seed Avg x Coverage	0.856	0.592	+44.6%
Betweenness	Seed Total	0.850	0.602	+41.2%
Complex Centrality (t = 2)	Seed Avg	0.845	0.889	–4.94%
Degree	Seed Avg x Coverage	0.843	0.580	+45.3%
Eigenvector	Seed Avg x Coverage	0.841	0.543	+54.9%
Influence Maximization (p = 0.10)	Seed Avg x Coverage	0.841	0.599	+40.4%
Influence Maximization (p = 0.01)	Seed Avg x Coverage	0.840	0.589	+42.6%
Betweenness	Seed Avg x Coverage	0.836	0.602	+38.9%
Influence Maximization (p = 0.50)	Seed Avg x Coverage	0.833	0.595	+40.0%
Complex Centrality (t = 4)	Seed Avg	0.804	0.825	–2.55%

Table 9: Top 10 Best Average Maximized Fair Interventions

We observe a change here in the prevailing fairness definition among the most effective maximized fair interventions, with *seed average* × *coverage fairness* now the most effective. This definition represents 7 out of the 10 best fair interventions that we could select, while *seed average fairness* and *seed total fairness* account for 2 and 1, respectively. The average improvement of these maximized fair interventions compared to their respective efficiency maximizing strategies is notable. Among the 8 out of 10 interventions that improve outcome fairness, the increase ranges from 38.9% to 54.9% improvement over the maximizing strategies. The actual average *outcome stat rates* for these interventions hold relatively high values as well, ranging from 0.833 to 0.856. These are relatively fair results.

Nevertheless, not all interventions follow this trend. Specifically, Complex Centrality t = 2 and t = 4 fail to augment outcome fairness relative to their corresponding maximizing strategies. Here, we notice losses of 4.94% (t = 2) and 2.55% (t = 4) in the *outcome stat rate* of their maximizing strategies. Although these reductions are not substantial, they become more pronounced when considering the overall impact on the spread of information (detailed more in the subsequent section). This observation highlights the lack of benefit in ensuring fairness for Complex Centrality within the specific networks used in this study. This observation does not categorically denote that Complex Centrality is unsuitable for fair interventions. However, assessing its capability in this domain is more challenging. Contrary to other network metrics, Complex Centrality was the only one where female villagers had dominant distributions. This aspect allowed it to yield substantially fair outcomes from straightforward maximization, which further complicates the evaluation of its effectiveness in fair interventions. That is, because it does not require a fair intervention.

The adoption of the *seed average* × *coverage fairness* definition significantly alters the success rates associated with maximized fair interventions. Employing this fairness constraint refines our strategic choices to those with an increased likelihood of achieving fair outcomes. Under this definition, Betweenness yields fair outcomes in 6 out of the 14 villages, as does Influence Maximization at all four parameter specifications (p = 0.01, 0.10, 0.25, and 0.50). Consequently, these network metrics collectively present an anticipated success rate of 42.28% if employed across all 14 villages. Both Eigenvector Centrality and Degree prove effective in 5 out of the 14 villages, translating to an effective success rate of 35.71%. These success rates include 7 of our 13 network metrics and are markedly higher than the 18.15% success rate of a randomly selected intervention. Alas, fairer outcomes might not prevail without certain trade-offs.

In the next section, we will explore the potential costs on the overall spread of information resulting from the implementation of fair interventions.

Costs of Fair Interventions on the Efficiency of Information Spread: Maximization strategies are commonly used interventions, as the typical goal for information diffusion is to optimize the efficiency of information spread, allowing it to reach the largest segment of the network as possible. Introducing fair interventions may consequently reduce the efficiency of information dissemination across the network because such interventions attempt to distribute influence equally instead of identifying the most influence, irrespective of groups.

Figure 5 presents the aggregate costs, expressed as a fraction of the maximized efficiency intervention, resulting from fair interventions in accordance with each network measure. These are density plots. Thus, the y-axis represents the count of villages (as cases), while the x-axis indicates the loss as a fraction of the corresponding maximizing strategy. The blue line represents the cost, as a fraction of each network measure's maximized efficiency intervention, borne from selecting the maximized fair intervention. Alternatively, the pink line highlights the cost, in terms of the fraction of each network measure's maximized efficiency intervention, incurred by strategies that resulted in the fairest outcome for that particular network measure and fairness definition. Within each network measure's plot are values for the average percentage loss for the maximized fair intervention (blue) and the fairest outcome (pink).

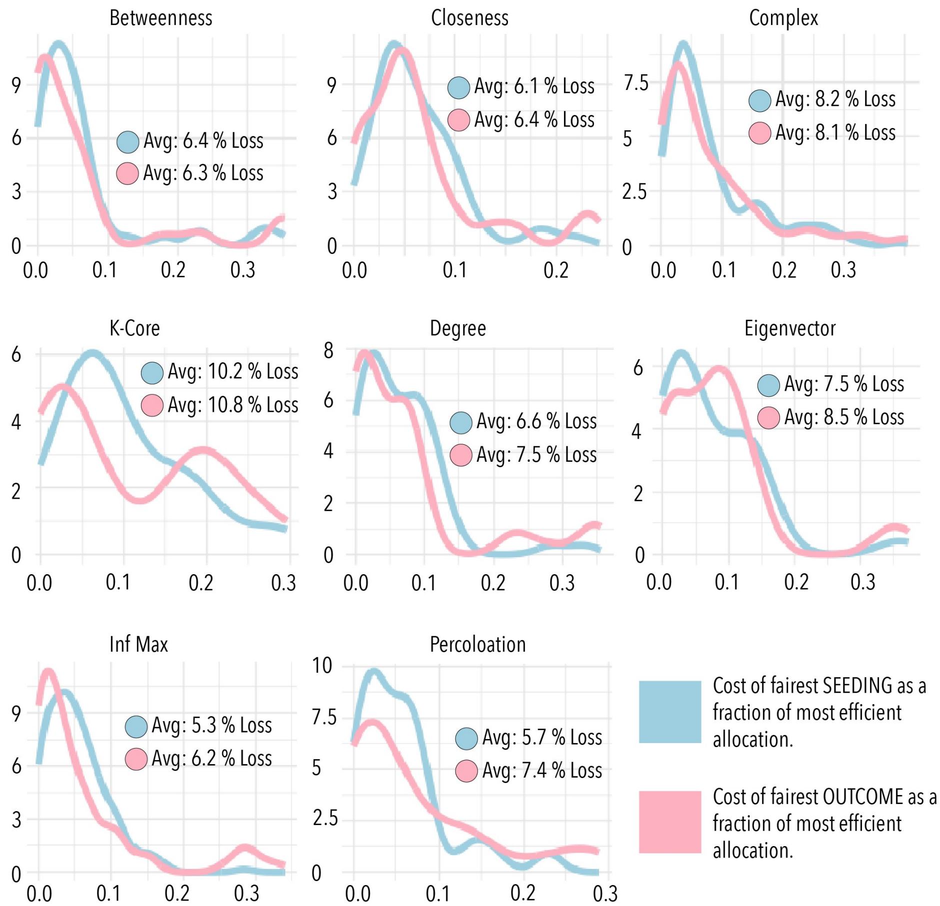

Figure 5: Costs in efficiency from fairness by strategy

The fair interventions exhibit diminished efficiency in comparison to their parallel maximization strategies across all examined network measures. There is no marked difference in the efficiency loss between maximized fair interventions and other interventions that produced the fairest outcomes, particularly for Degree, Betweenness, and Complex Centrality measures where the differences are negligible. An exception is Eigenvector Centrality, where a significant portion of the villages experience more substantial losses for the interventions that produce the fairest outcomes (about 0.1) compared to maximized fair interventions (about 0.05). Despite slight variations, the average loss remains slightly higher for interventions that produce the fairest outcome interventions, with the exception of Betweenness and Complex Centrality. Among the most effective strategies for generating fair outcomes — Complex Centrality, Betweenness, Influence Maximization, Degree, and Eigenvector — the incurred losses are relatively uniform, typically ranging between 0 and 10%.

Influence Maximization and Betweenness register the lowest average losses, albeit not significantly different from Degree, Eigenvector, and Complex Centrality.

Table 10 outlines the percentage loss in aggregate information spread, categorized by village and calculated as a percentage of the average maximized efficiency intervention — e.g., the percentage cost of not selecting the maximized efficiency intervention. The second column lists the average percentage loss of fair interventions in each village compared to their corresponding maximized efficiency interventions. The third column displays the average loss for interventions that realized the fairest outcomes for each network measure, expressed as a percentage of each network measure's maximized efficiency intervention for each village. The fourth column presents average loss for the set of maximized fair interventions that received *outcome stat rates* ≥ 0.9 relative to average maximized efficiency interventions for each village.

Variations in losses between fair and maximizing strategies are more accentuated when observed across different villages. This observation implies that diverse network settings yield disparate responses to fair seeding. Such differences can either amplify or diminish the disparity in information spread between interventions focused on maximized efficiency and those prioritizing fairness.

In 9 out of the 14 villages, the losses are higher for fair interventions compared to those interventions that yield the fairest outcomes. The range of values is more extensive across villages than the prior observed losses based on network metrics. Values across villages vary from 4.9% to 11.9% for fair strategies, and from 0.6% to 24.2% for interventions that resulted in the fairest outcomes.

The fair interventions (column 2) and interventions yielding the fairest outcomes (column 3) are not necessarily selectable interventions. They merely represent average figures for fair interventions and outcomes. In column 4, the losses of maximized fair interventions are averaged as percentages of the corresponding maximizing strategies, representing losses associated with possible selectable strategies. This analysis reveals a broader range of losses, stretching from 0.6% to 49.9%. Some values are absent, indicating villages where achieving fair outcomes with a maximized fair intervention was unattainable. Particularly large losses, like

Village	Avg. FS* vs Avg. Max (% Loss)	Avg. FO* vs Avg. Max (% Loss)	Avg. Fairest vs Avg. Max (% Loss)
1	9.9%	7.8%	9.7%
2	11.2%	9.3%	6.7%
3	9.7%	5.4%	39.6%
4	7.7%	2.7%	8.0%
5	5.1%	2.5%	-
6	7.8%	5.8%	18.7%
7	6.4%	10.4%	-
8	7.0%	4.2%	13.7%
9	11.8%	18.1%	-
10	10.2%	0.9%	0.6%
11	11.9%	24.2%	49.9%
12	4.9%	5.2%	6.2%
13	5.4%	1.6%	5.9%
14	10.3%	7.0%	10.2%

* FS: Fair Seeding   * FO: Fair Outcome

Table 10: Costs in efficiency from fairness by village

in village 3 (39.6%) and village 11 (49.9%), are noteworthy as they indicate failures in information spread under fair interventions.

In general, losses across fair interventions relative to corresponding efficiency maximizing strategies show some consistency. However, the significance of these losses requires a contextual understanding. Determining whether a 5% decrease in information spread is a notable cost is not straightforward. More specifically, it is challenging to ascertain whether this loss significantly undermines the pursuit of equitable outcomes. This evaluation inherently depends on individual perspectives regarding the costs associated with unequal outcomes. In certain contexts, like disease treatment, both efficiency and fairness hold substantial importance, posing ethical dilemmas. Conversely, in situations like specific marketing scenarios, decisions regarding the trade-off between efficiency and fairness may be easier to reconcile.

4 Discussion and Conclusion

This study began by spotlighting information diffusion processes as potential systems through which group inequalities become exacerbated. Conventional methodologies for initializing diffusion processes, particularly those emphasizing the maximization of efficiency, can inadvertently amplify disparities across divergent groups. Certain advantaged groups may be endowed with beneficial network characteristics and information gathering privileges, generating biases in network interventions and the downstream outcomes of diffusion processes. A framework for the analysis of fairness, extracted from the field of computation and algorithmic fairness, has been employed as an analytical tool for understanding and evaluating disparities resulting from information diffusion processes.

Our examination scrutinized a large set of common network metrics for influence and demonstrated how naive seeding of maximizing strategies would produce unequally represented seed sets across male and female villagers using our case of 14 independent Ugandan villages. Utilizing a sophisticated Simulated Method of Moments approach, a diffusion model grounded in the 14 actual, independent diffusion processes was used to explore the relationship between fairness in interventions and fairness in diffusion outcomes. We tested four foundational fairness definitions that ensured the equitable distribution of network influence in the seeding of information diffusion processes based on the seed set or the village population. An extensive set of simulation tests, encompassing 13 prevalent network metrics, was deployed to examine the efficacy of interventions under network fairness constraints in fostering equitable outcomes in information diffusion.

Our initial empirical findings underscored that the likelihood of a randomly chosen fair intervention resulting in a fair outcome is demonstrably lower compared to the random selection of an intervention from the set of all possible allocations. A granular assessment by network measure identified Betweenness as a slight improvement over purely random selection for yielding fair outcomes. Despite these observations, maximized fair interventions, which denote selectable interventions, exhibited even lower success rates for generating fair outcomes. Even so, we found fair interventions produced a general improvement in outcome fairness over corresponding maximizing strategies. Our subsequent introduction of a new fairness definition, *seed average x coverage fairness*, improved the prediction of outcome fairness across a much larger set of network metrics than our four original definitions. We observed under *seed average x coverage fairness* increased success rates for our maximized fair interventions such that they were more consistent predictors of fair outcomes than selecting a random intervention. We also observed significant improvements to average level of outcome fairness for 7 of our 13 network metrics. Nonetheless, we observed consistent losses in the efficiency of spreading information when choosing fair interventions over corresponding maximizing strategies. These losses raise a ethical question about the trade-offs between efficiency and fairness in diffusion settings, though we consider such conflicts to be context dependent.

Our study faced several limitations. Predominantly empirical, the exploration spanned a limited array of real cases and did not aim to provide explanations that were generalized to the performance of fair interventions in association with specific network characteristics. While employing actual cases imparts realistic aspects and depth to our examination of the relationship between fair interventions and fair outcomes, the same case-specific granularity simultaneously limits the broad applicability of our insights. This inherent limitation points towards potential future work, where exploration of network topology could emerge as an important focus, providing a richer understanding and broader applicability in understanding the dynamics of fair interventions in network settings.

In this exploration, our analysis is centered on straightforward fairness definitions, proposing these as fundamental benchmarks for evaluating the fairness of interventions. Such a focus allows for a clearer, albeit basic, examination of the interplay between fairness and intervention strategies within network settings.
Future work could potentially expand this scope by investigating a more extensive set of fairness definitions. Additionally, a broader exploration of seeding strategies, network measures, or alternative metrics for influence could enrich the understanding and implementation of fairness in network interventions, offering a more comprehensive and nuanced view of the dynamics involved.

The models utilized were calibrated based on a set of independent real cases, yet they still represent a somewhat simplified portrayal of the diffusion processes. A logical progression from this research would be to examine fair interventions within actual diffusion processes to gauge their real-world effectiveness and implications. The insights gained from such practical, real-world testing could significantly enhance the depth and applicability of our modeled research, providing a deeper understanding of fair interventions in diffusion processes.

References

Acker, J. (1990). Hierarchies, Jobs, Bodies: A Theory of Gendered Organizations [Publisher: SAGE Publications Inc]. *Gender & Society*, *4*(2), 139–158. https://doi.org/10.1177/089124390004002002

Ali, J., Babaei, M., Chakraborty, A., Mirzasoleiman, B., Gummadi, K. P., & Singla, A. (2019). On the Fairness of Time-Critical Influence Maximization in Social Networks [eprint: 1905.06618].

Amirkhanian, Y. A., Kelly, J. A., & McAuliffe, T. L. (2005). Identifying, recruiting, and assessing social networks at high risk for HIV/AIDS: Methodology, practice, and a case study in St Petersburg, Russia. *AIDS care*, *17*(1), 58–75. https://doi.org/10.1080/09540120412331305133

Aral, S., & Van Alstyne, M. (2011). The Diversity-Bandwidth Trade-off. *American Journal of Sociology*, *117*(1), 90–171. https://doi.org/10.1086/661238

Aral, S., & Walker, D. (2011). Creating Social Contagion Through Viral Product Design: A Randomized Trial of Peer Influence in Networks [Publisher: INFORMS]. *Management Science*, *57*(9), 1623–1639. https://doi.org/10.1287/mnsc.1110.1421

Arslan, C., Wollni, M., Oduol, J., & Hughes, K. (2022). Who communicates the information matters for technology adoption. *World Development*, *158*, 106015. https://doi.org/10.1016/j.worlddev.2022. 106015

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013). The Diffusion of Microfinance [Publisher: American Association for the Advancement of Science]. *Science*, *341*(6144), 1236498. https://doi.org/10.1126/science.1236498

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2014). Gossip: Identifying Central Individuals in a Social Network. https://doi.org/10.3386/w20422

Barbeau, E. M., Wolin, K. Y., Naumova, E. N., & Balbach, E. (2005). Tobacco advertising in communities: Associations with race and class. *Preventive Medicine*, *40*(1), 16–22. https://doi.org/10.1016/j.ypmed.2004.04.056

Beaman, L., & Dillon, A. (2018). Diffusion of agricultural information within social networks: Evidence on gender inequalities from Mali. *Journal of Development Economics*, *133*, 147–161. https://doi.org/10.1016/j.jdeveco.2018.01.009

Benjamin, R. (2019a). *Captivating Technology: Race, Carceral Technoscience, and Liberatory Imagination in Everyday Life* [Google-Books-ID: aOiZDwAAQBAJ]. Duke University Press.

Benjamin, R. (2019b). Assessing risk, automating racism [Publisher: American Association for the Advancement of Science]. *Science*, *366*(6464), 421–422. https://doi.org/10.1126/science.aaz3873

Berk, R., Heidari, H., Jabbari, S., Joseph, M., Kearns, M. J., Morgenstern, J., Neel, S., & Roth, A. (2017). A Convex Framework for Fair Regression. *CoRR*, *abs/1706.02409*.

Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2017). Fairness in Criminal Justice Risk Assessments: The State of the Art. *arXiv:1703.09207 [stat]*. Retrieved May 4, 2020, from http://arxiv.org/abs/1703.09207

Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2018). Fairness in criminal justice risk assessments: The state of the art [Publisher: Sage Publications Sage CA: Los Angeles, CA]. *Sociological Methods & Research*, 0049124118782533.

Bonacich, P. (1987). Power and Centrality: A Family of Measures. *American Journal of Sociology*, *92*(5), 1170–1182. http://www.jstor.org/stable/2780000

Browne, S. (2015). *Dark Matters: On the Surveillance of Blackness* [Google-Books-ID: snmJCgAAQBAJ]. Duke University Press.

Burt, R. S. (1995). *Structural Holes: The Social Structure of Competition* (1st Paperback Edition). Harvard University Press.

Burt, R. S. (1999). The Social Capital of Opinion Leaders [Publisher: SAGE Publications Inc]. *The ANNALS of the American Academy of Political and Social Science*, *566*(1), 37–54. https://doi.org/10.1177/000271629956600104

Burt, R. S. (2007). *Brokerage and Closure: An Introduction to Social Capital*. Oxford University Press.

Campos-Castillo, C., & Laestadius, L. I. (2020). Racial and Ethnic Digital Divides in Posting COVID-19 Content on Social Media Among US Adults: Secondary Survey Analysis. *Journal of Medical Internet Research*, *22*(7), e20472. https://doi.org/10.2196/20472

Carmel, D., Zwerdling, N., Guy, I., Ofek-Koifman, S., Har'El, N., Ronen, I., Uziel, E., Yogev, S., & Chernov, S. (2009). Personalized social search based on the user's social network. *Proceedings of the 18th ACM conference on Information and knowledge management*, 1227–1236.

Celis, L. E., Mehrotra, A., & Vishnoi, N. K. (2019). Toward Controlling Discrimination in Online Ad Auctions. *ICML*, *97*, 4456–4465.

Centola, D. (2011). An experimental study of homophily in the adoption of health behavior. *Science*. https://doi.org/10.1126/science.1207055

Centola, D. (2015). The Social Origins of Networks and Diffusion. *American Journal of Sociology*, *120*(5), 1295–1338. https://doi.org/10.1086/681275

Centola, D., Egu´ıluz, V. M., & Macy, M. W. (2007). Cascade dynamics of complex propagation. *Physica A: Statistical Mechanics and its Applications*, *374*(1), 449–456. https://doi.org/10.1016/j.physa.2006. 06.018

Centola, D., & Macy, M. (2007). Complex Contagions and the Weakness of Long Ties. *American Journal of Sociology*, *113*(3), 702–734. https://doi.org/10.1086/521848

Centola, D., & van de Rijt, A. (2015). Choosing your network: Social preferences in an online health community. *Social Science & Medicine*, *125*, 19–31. https://doi.org/10.1016/j.socscimed.2014.05.019 Chatterjee, P. (2011). Drivers of new product recommending and referral behaviour on social network sites [Publisher: Taylor & Francis]. *International Journal of Advertising*, *30*(1), 77–101.

Chen, J., Kallus, N., Mao, X., Svacha, G., & Udell, M. (2019). Fairness Under Unawareness: Assessing Disparity When Protected Class Is Unobserved. *FAT*, 339–348.

Chen, X., Fain, B., Lyu, C., & Munagala, K. (2019). Proportionally Fair Clustering. *The 36th International Conference on Machine Learning (ICML)*.

Cheng, H. W. J. (2022). Factors Affecting Technological Diffusion Through Social Networks: A Review of the Empirical Evidence. *The World Bank Research Observer*, *37*(2), 137–170. https://doi.org/10.1093/wbro/lkab010

Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., Gong, S., Gonzalez, F., Grondin, A., Jacob, M., Johnston, D., Koenen, M., Laguna-Muggenburg, E., Mudekereza, F., Rutter, T., Thor, N., Townsend, W., Zhang, R., Bailey, M., ... Wernerfelt, N. (2022). Social capital II: Determinants of economic connectedness [Number: 7921 Publisher: Nature Publishing Group]. *Nature*, *608*(7921), 122–134. https://doi.org/10.1038/s41586-022-04997-3

Chouldechova, A. (2016). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *arXiv:1610.07524 [cs, stat]*. Retrieved May 4, 2020, from http://arxiv.org/abs/1610.07524

Chouldechova, A., Benavides-Prado, D., Fialko, O., & Vaithianathan, R. (2018). A case study of algorithmassisted decision making in child maltreatment hotline screening decisions [ISSN: 2640-3498]. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 134–148. Retrieved September 4, 2023, from https://proceedings.mlr.press/v81/chouldechova18a.html

Christakis, N. A., & Fowler, J. H. (2013). Social contagion theory: Examining dynamic social networks and human behavior [Publisher: Wiley Online Library]. *Statistics in medicine*, *32*(4), 556–577.

Chu, K.-H., Matheny, S., Furek, A., Sidani, J., Radio, S., Miller, E., Valente, T., & Robertson, L. (2021). Identifying student opinion leaders to lead e-cigarette interventions: Protocol for a randomized controlled pragmatic trial. *Trials*, *22*(1), 31. https://doi.org/10.1186/s13063-020-04990-z

Cobb, N. K., Jacobs, M. A., Wileyto, P., Valente, T., & Graham, A. L. (2016). Diffusion of an evidencebased smoking cessation intervention through Facebook: A randomized controlled trial [Publisher: American Public Health Association]. *American journal of public health*, *106*(6), 1130–1135.

Coleman, J. (1988). Social Capital in the Creation of Human Capital. *The American Journal of Sociology*, *94*, S95–S120. http://www.jstor.org/stable/2780243

Coleman, J., Katz, E., & Menzel, H. (1957). The Diffusion of an Innovation Among Physicians [Publisher: [American Sociological Association, Sage Publications, Inc.]]. *Sociometry*, *20*(4), 253–270. https://doi.org/10.2307/2785979

Coleman, J., Katz, E., & Menzel, H. (1966). *Medical Innovation: A Diffusion Study* (First Printing edition). Bobbs-Merrill Co.

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017). Algorithmic Decision Making and the Cost of Fairness. *KDD*, 797–806.

Coulter, P. B. (2019). *Measuring Inequality: A Methodological Handbook* [Google-Books-ID: dcPADwAAQBAJ]. Routledge.

Cowell, F. (2011). *Measuring Inequality* (3rd ed. edition). Oxford University Press.

Cross, R., & Sproull, L. (2004). More Than an Answer: Information Relationships for Actionable Knowledge. *Organization Science*, *15*(4), 446–462. https://doi.org/10.1287/orsc.1040.0075

DiMaggio, P., & Garip, F. (2011). How Network Externalities Can Exacerbate Intergroup Inequality. *American Journal of Sociology*, *116*(6), 1887–1933. https://doi.org/10.1086/659653

DiMaggio, P., & Garip, F. (2012). Network Effects and Social Inequality. *Annual Review of Sociology*, *38*(1), 93–118. https://doi.org/10.1146/annurev.soc.012809.102545

DiPrete, T. A., Gelman, A., McCormick, T., Teitler, J., & Zheng, T. (2011). Segregation in Social Networks Based on Acquaintanceship and Trust [Publisher: The University of Chicago Press]. *American Journal of Sociology*, *116*(4), 1234–83. https://doi.org/10.1086/659100

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. S. (2012). Fairness through awareness. In S. Goldwasser (Ed.), *Innovations in Theoretical Computer Science 2012, Cambridge, MA, USA, January 8-10, 2012* (pp. 214–226). ACM.

Eagle, N., Pentland, A. (, & Lazer, D. (2009). Inferring friendship network structure by using mobile phone data [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *106*(36), 15274–15278. https://doi.org/10.1073/pnas.0900282106

Easley, D., & Kleinberg, J. (2010). *Networks, Crowds, and Markets: Reasoning about a Highly Connected World* (1st edition). Cambridge University Press.

Edwards, K. M., Banyard, V. L., Waterman, E. A., Hopfauf, S. L., Shin, H.-S., Simon, B., & Valente, T. W. (2022). Use of Social Network Analysis to Identify Popular Opinion Leaders for a Youth-Led Sexual Violence Prevention Initiative. *Violence Against Women*, *28*(2), 664–685. https://doi.org/10.1177/1077801221994907

Elliott, J. R., & Smith, R. A. (2004). Race, Gender, and Workplace Power [Publisher: [American Sociological Association, Sage Publications, Inc.]]. *American Sociological Review*, *69*(3), 365–386. Retrieved September 8, 2023, from https://www.jstor.org/stable/3593052

Eubanks, V. (2011). *Digital Dead End: Fighting for Social Justice in the Information Age*. The MIT Press. Retrieved September 27, 2023, from https://www.jstor.org/stable/j.ctt5hhgk1

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press.

Evans, R. (2018). Simulated Method of Moments (SMM) Estimation. https : / / notes . quantecon . org / submission/5b3db2ceb9eab00015b89f93

Evans, R. (2023). Simulated Method of Moments Estimation. In *Computational Methods for Economists using Python* (v0.0.2). Open access Jupyter Book. https://opensourceecon.github.io/CompMethods/ struct est/SMM.html

Ferrali, R., Grossman, G., Platas, M. R., & Rodden, J. (2020). It takes a village: Peer effects and externalities in technology adoption [Publisher: Wiley Online Library]. *American Journal of Political Science*, *64*(3), 536–553.

Fish, B., Bashardoust, A., boyd danah, d., Friedler, S. A., Scheidegger, C., & Venkatasubramanian, S. (2019). Gaps in Information Access in Social Networks? In L. Liu, R. W. White, A. Mantrach, F. Silvestri, J. J. McAuley, R. Baeza-Yates, & L. Zia (Eds.), *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019* (pp. 480–490). ACM.

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter Bubbles, Echo Chambers, and Online News Consumption [eprint: /oup/backfile/content public/journal/poq/80/s1/10.1093 poq nfw006/4/nfw006.pdf]. *Public Opinion Quarterly*, *80*(S1), 298–320. https://doi.org/10.1093/poq/nfw006

Foulds, J. R., Islam, R., Keya, K. N., & Pan, S. (2020). An Intersectional Definition of Fairness [ISSN: 2375-026X]. *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, 1918–1921. https://doi.org/10.1109/ICDE48307.2020.00203

Freeman, L. C. (1977). A Set of Measures of Centrality Based on Betweenness. *Sociometry*, *40*(1), 35–41. https://doi.org/10.2307/3033543

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239. https://doi.org/10.1016/0378-8733(78)90021-7

Freeman, L. C., Roeder, D., & Mulholland, R. R. (1979). Centrality in social networks: II. Experimental results. *Social Networks*, *2*(2), 119–141. https://doi.org/10.1016/0378-8733(79)90002-9

Gauthier, G. R., Smith, J. A., Garc´ıa, C., Garcia, M. A., & Thomas, P. A. (2020). Exacerbating Inequalities: Social Networks, Racial/Ethnic Disparities, and the COVID-19 Pandemic in the United States. *J Gerontol B Psychol Sci Soc Sci*. https://doi.org/10.1093/geronb/gbaa117

Godes, D., & Mayzlin, D. (2004). Using Online Conversations to Study Word-of-Mouth Communication [Publisher: INFORMS]. *Marketing Science*, *23*(4), 545–560. Retrieved September 6, 2023, from https://www.jstor.org/stable/30036688

Goldenberg, J., Han, S., Lehmann, D. R., & Hong, J. W. (2009). The Role of Hubs in the Adoption Process [Publisher: American Marketing Association]. *Journal of Marketing*, *73*(2), 1–13. Retrieved September 6, 2023, from https://www.jstor.org/stable/20619007

Granovetter, M. (1973). The Strength of Weak Ties. *American Journal of Sociology*, *78*(6), 1360–1380.

Granovetter, M. (1978). Threshold models of collective behavior [Publisher: JSTOR]. *American Journal of Sociology*, 1420–1443.

Graves, S. M. (2003). Landscapes of Predation, Landscapes of Neglect: A Location Analysis of Payday Lenders and Banks. *The Professional Geographer*, *55*(3), 303–317. https://doi.org/10.1111/00330124.5503017

Guilbeault, D., & Centola, D. (2021). Topological measures for identifying and predicting the spread of complex contagions [Bandiera abtest: a Cc license type: cc by Cg type: Nature Research Journals Number: 1 Primary atype: Research Publisher: Nature Publishing Group Subject term: Society;Sociology Subject term id: society;sociology]. *Nature Communications*, *12*(1), 4430. https://doi.org/10.1038/ s41467-021-24704-6

Hanaki, N., Peterhansl, A., Dodds, P. S., & Watts, D. J. (2007). Cooperation in Evolving Social Networks [Publisher: INFORMS]. *Management Science*, *53*(7), 1036–1050. Retrieved September 6, 2023, from https://www.jstor.org/stable/20122268

Hansen, M. (1999). The Search-Tranfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits. *Administrative Science Quarterly*, *44*(1), 82–111.

Hanson, W. A., & Putler, D. S. (1996). Hits and misses: Herd behavior and online product popularity. *Marketing Letters*, *7*(4), 297–305. Retrieved April 11, 2015, from http://link.springer.com/article/10.1007/BF00435537

Harder, B. M., & Sumerau, J. E. (2018). Understanding Gender as a Fundamental Cause of Health: Simultaneous Linear Relationships between Gender, Mental Health, and Physical Health Over Time. *Sociological Spectrum*, *38*(6), 387–405. https://doi.org/10.1080/02732173.2018.1532366

Hardt, M., Price, E., & Srebro, N. (2016). Equality of Opportunity in Supervised Learning. In D. D. Lee, M. Sugiyama, U. v. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain* (pp. 3315–3323). https://proceedings.neurips.cc/paper/2016/hash/9d2682367c3935defcb1f9e247a97c0d-Abstract.html

Hinton, E., & Cook, D. (2021). The Mass Criminalization of Black Americans: A Historical Overview [eprint: https://doi.org/10.1146/annurev-criminol-060520-033306]. *Annual Review of Criminology*, *4*(1), 261–286. https://doi.org/10.1146/annurev-criminol-060520-033306

Hinz, O., Skiera, B., Barrot, C., & Becker, J. U. (2011). Seeding Strategies for Viral Marketing: An Empirical Comparison [Publisher: SAGE Publications Inc]. *Journal of Marketing*, *75*(6), 55–71. https://doi.org/10.1509/jm.10.0088

Ibarra, H. (1992). Homophily and Differential Returns: Sex Differences in Network Structure and Access in an Advertising Firm. *Administrative Science Quarterly*, *37*(3), 422–447. https://doi.org/10.2307/2393451

Ioannides, Y. M., & Loury, L. D. (2004). Job Information Networks, Neighborhood Effects, and Inequality. *Journal of Economic Literature*, *42*(4), 1056–1093. https://doi.org/10.1257/0022051043004595

Iyengar, R., Van den Bulte, C., & Lee, J. Y. (2015). Social Contagion in New Product Trial and Repeat [Publisher: INFORMS]. *Marketing Science*, *34*(3), 408–429. Retrieved September 5, 2023, from https://www.jstor.org/stable/24544841

Iyengar, R., Van den Bulte, C., & Valente, T. W. (2011). Opinion Leadership and Social Contagion in New Product Diffusion [Publisher: INFORMS]. *Marketing Science*, *30*(2), 195–212. https://doi.org/10.1287/mksc.1100.0566

Kalev, A. (2009). Cracking the Glass Cages? Restructuring and Ascriptive Inequality at Work [Publisher: The University of Chicago Press]. *American Journal of Sociology*, *114*(6), 1591–1643. https://doi.org/10.1086/597175

Kanter, R. M. (1993). *Men and Women of the Corporation: New Edition* (New edition). Basic Books.

Kelly, J. A., Heckman, T. G., Stevenson, L. Y., Williams, P. N., Ertl, T., Hays, R. B., Leonard, N. R., O'Donnell, L., Terry, M. A., Sogolow, E. D., & Neumann, M. S. (2000). Transfer of research-based HIV prevention interventions to community service providers: Fidelity and adaptation. *AIDS education and prevention: official publication of the International Society for AIDS Education*, *12*(5 Suppl), 87–98.

Kelly, J. A., Murphy, D. A., Sikkema, K. J., McAuliffe, T. L., Roffman, R. A., Solomon, L. J., Winett, R. A., & Kalichman, S. C. (1997). Randomised, controlled, community-level HIV-prevention intervention for sexual-risk behaviour among homosexual men in US cities. Community HIV Prevention Research Collaborative. *Lancet (London, England)*, *350*(9090), 1500–1505. https://doi.org/10.1016/s01406736(97)07439-4

Kelly, J. A., Amirkhanian, Y. A., Kabakchieva, E., Vassileva, S., Vassilev, B., McAuliffe, T. L., DiFranceisco, W. J., Antonova, R., Petrova, E., Khoursine, R. A., & Dimitrov, B. (2006). Prevention of HIV and sexually transmitted diseases in high risk social networks of young Roma (Gypsy) men in Bulgaria: Randomised controlled trial [Publisher: British Medical Journal Publishing Group Section: Research]. *BMJ*, *333*(7578), 1098. https://doi.org/10.1136/bmj.38992.478299.55

Kempe, D., Kleinberg, J., & Tardos, E. (2003). Maximizing the spread of influence through a social network.´ *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 137–146.

Kim, D. A., Hwong, A. R., Stafford, D., Hughes, D. A., O'Malley, A. J., Fowler, J. H., & Christakis, N. A. (2015). Social network targeting to maximise population behaviour change: A cluster randomised controlled trial [Publisher: Elsevier]. *The Lancet*, *386*(9989), 145–153.

Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., & Makse, H. A. (2010). Identification of influential spreaders in complex networks [Number: 11 Publisher: Nature Publishing Group]. *Nature Physics*, *6*(11), 888–893. https://doi.org/10.1038/nphys1746

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2016). Inherent Trade-Offs in the Fair Determination of Risk Scores [arXiv: 1609.05807]. *arXiv:1609.05807 [cs, stat]*. Retrieved March 8, 2022, from http://arxiv.org/abs/1609.05807

Kozodoi, N., Jacob, J., & Lessmann, S. (2022). Fairness in credit scoring: Assessment, implementation and profit implications. *European Journal of Operational Research*, *297*(3), 1083–1094. https://doi.org/10.1016/j.ejor.2021.06.023

Lambrecht, A., & Tucker, C. E. (2018). *Algorithmic Bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads* (SSRN Scholarly Paper No. ID 2852260). Social Science Research Network. Rochester, NY. https://doi.org/10.2139/ssrn.2852260

Larson, J., Mattu, S., Kirchner, L., & Angwin, J. (2016). How we analyzed the COMPAS recidivism algorithm. *ProPublica (5 2016)*, *9*.

Lehmann, S., & Ahn, Y.-Y. (Eds.). (2019). *Complex Spreading Phenomena in Social Systems: Influence and Contagion in Real-World Social Networks* (Softcover reprint of the original 1st ed. 2018 edition). Springer.

Leskovec, J., Adamic, L. A., & Huberman, B. A. (2007). The dynamics of viral marketing. *ACM Transactions on the Web*, *1*(1), 5–es. https://doi.org/10.1145/1232722.1232727

Levin, D. Z., & Cross, R. (2004). The Strength of Weak Ties You Can Trust: The Mediating Role of Trust in Effective Knowledge Transfer [Publisher: INFORMS]. *Management Science*, *50*(11), 1477–1490. https://doi.org/10.1287/mnsc.1030.0136

Li, M., Wang, X., Gao, K., & Zhang, S. (2017). A Survey on Information Diffusion in Online Social Networks: Models and Methods [Number: 4 Publisher: Multidisciplinary Digital Publishing Institute]. *Information*, *8*(4), 118. https://doi.org/10.3390/info8040118

Li, Y., Fan, J., Wang, Y., & Tan, K.-L. (2018). Influence Maximization on Social Graphs: A Survey [Conference Name: IEEE Transactions on Knowledge and Data Engineering]. *IEEE Transactions on Knowledge and Data Engineering*, *30*(10), 1852–1872. https://doi.org/10.1109/TKDE.2018.2807843

Link, B., & Phelan, J. (1995). Social Conditions As Fundamental Causes of Disease. *Journal of Health and Social Behavior*, 80–94. https://doi.org/10.2307/2626958

Link, B., & Phelan, J. (2010). Social Conditions as Fundamental Causes of Health Inequalities. In C. Bird, P. Conrad, A. M. Fremont, & S. Timmermans (Eds.), *Handbook of Medical Sociology, Sixth Edition*. Vanderbilt University Press.

Lungeanu, A., McKnight, M., Negron, R., Munar, W., Christakis, N. A., & Contractor, N. S. (2021). Using Trellis software to enhance high-quality large-scale network data collection in the field. *Social Networks*, *66*, 171–184. https://doi.org/10.1016/j.socnet.2021.02.007

Lutfey, K., & Freese, J. (2005). Toward Some Fundamentals of Fundamental Causality: Socioeconomic Status and Health in the Routine Clinic Visit for Diabetes. *American Journal of Sociology*, *110*(5), 1326– 1372. https://doi.org/10.1086/428914

Marsden, P. V. (1987). Core Discussion Networks of Americans. *American Sociological Review*, *52*(1), 122– 131. https://doi.org/10.2307/2095397

Marsden, P. V. (1988). Homogeneity in confiding relations. *Social Networks*, *10*(1), 57–76. https://doi.org/10.1016/0378-8733(88)90010-X

Maume Jr., D. J. (1999). Glass ceilings and glass escalators: Occupational segregation and race and sex differences in managerial promotions [Place: US Publisher: Sage Publications]. *Work and Occupations*, *26*(4), 483–509. https://doi.org/10.1177/0730888499026004005

McPherson, M., Smith-Lovin, L., & Brashears, M. E. (2006). Social Isolation in America: Changes in Core Discussion Networks over Two Decades [Publisher: SAGE Publications Inc]. *American Sociological Review*, *71*(3), 353–375. https://doi.org/10.1177/000312240607100301

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks [Publisher: Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA]. *Annual review of sociology*, *27*(1), 415–444.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2022). A Survey on Bias and Fairness in Machine Learning [arXiv:1908.09635 [cs]]. https://doi.org/10.48550/arXiv.1908.09635

Mehrotra, A., Sachs, J., & Celis, L. E. (2022). Revisiting Group Fairness Metrics: The Effect of Networks [Place: New York, NY, USA Publisher: Association for Computing Machinery]. *Proc. ACM Hum.Comput. Interact.*, *6*(CSCW2). https://doi.org/10.1145/3555100

Morone, F., & Makse, H. A. (2015). Influence maximization in complex networks through optimal percolation [Number: 7563 Publisher: Nature Publishing Group]. *Nature*, *524*(7563), 65–68. https://doi.org/10.1038/nature14604

Muller, C., & Schrage, D. (2014). Mass Imprisonment and Trust in the Law [Publisher: SAGE Publications Inc]. *The ANNALS of the American Academy of Political and Social Science*, *651*(1), 139–158. https://doi.org/10.1177/0002716213502928

Narayanan, A. (2018). Tutorial: 21 Fairness Definitions and Their Politics. https://www.youtube.com/watch?v=jIXIuYdnyyk

Nelson, A. (2016). *The Social Life of DNA: Race, Reparations, and Reconciliation After the Genome* [GoogleBooks-ID: Xd7YCwAAQBAJ]. Beacon Press.

Nematzadeh, A., Ferrara, E., Flammini, A., & Ahn, Y.-Y. (2014). Optimal Network Modularity for Information Diffusion [Publisher: American Physical Society]. *Physical Review Letters*, *113*(8), 088701. https://doi.org/10.1103/PhysRevLett.113.088701

Newman, M. (2018). *Networks* (2nd edition). Oxford University Press.

Padgett, J. F., & Ansell, C. K. (1993). Robust Action and the Rise of the Medici, 1400-1434. *American Journal of Sociology*, *98*(6), 1259–1319.

Pager, D., & Pedulla, D. S. (2015). Race, Self-Selection, and the Job Search Process [Publisher: The University of Chicago Press]. *American Journal of Sociology*, *120*(4), 1005–1054. https://doi.org/10.1086/ 681072

Pager, D., & Shepherd, H. (2008). The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets. *Annual review of sociology*, *34*, 181.

Pedulla, D., & Pager, D. (2019). Race and Networks in the Job Search Process. *American Sociological Review*, *84*, 000312241988325. https://doi.org/10.1177/0003122419883255

Peres, R., Muller, E., & Mahajan, V. (2010). Innovation diffusion and new product growth models: A critical review and research directions. *International Journal of Research in Marketing*, *27*(2), 91–106. https://doi.org/10.1016/j.ijresmar.2009.12.012

Pessach, D., & Shmueli, E. (2022). A Review on Fairness in Machine Learning. *ACM Computing Surveys*, *55*(3), 51:1–51:44. https://doi.org/10.1145/3494672

Petersen, T., Saporta, I., & Seidel, M.-D. L. (2000). Offering a Job: Meritocracy and Social Networks [Publisher: The University of Chicago Press]. *American Journal of Sociology*, *106*(3), 763–816. https://doi.org/10.1086/318961

Phelan, J., & Link, B. (2015). Is Racism a Fundamental Cause of Inequalities in Health? [eprint: https://doi.org/10.1146/annu soc-073014-112305]. *Annual Review of Sociology*, *41*(1), 311–330. https://doi.org/10.1146/annurevsoc-073014-112305

Pickering, T. A., Wyman, P. A., Schmeelk-Cone, K., Hartley, C., Valente, T. W., Pisani, A. R., Rulison, K. L., Brown, C. H., & LoMurray, M. (2018). Diffusion of a Peer-Led Suicide Preventive Intervention Through School-Based Student Peer and Adult Networks. *Frontiers in Psychiatry*, *9*, 598. https://doi.org/10.3389/fpsyt.2018.00598

Pleiss, G., Raghavan, M., Wu, F., Kleinberg, J. M., & Weinberger, K. Q. (2017). On Fairness and Calibration. *NIPS*, 5680–5689.

Porter, M. A., & Gleeson, J. P. (2016). *Dynamical Systems on Networks: A Tutorial*. Springer International Publishing. https://doi.org/10.1007/978-3-319-26641-1

Rahmattalabi, A., Vayanos, P., Fulginiti, A., Rice, E., Wilder, B., Yadav, A., & Tambe, M. (2020). Exploring Algorithmic Fairness in Robust Graph Covering Problems. *arXiv:2006.06865 [cs, math]*. Retrieved January 20, 2021, from http://arxiv.org/abs/2006.06865

Rawls, J. (1999). *A Theory of Justice* (2nd edition). Belknap Press: An Imprint of Harvard University Press.

Reagans, R., & McEvily, B. (2003). Network structure and knowledge transfer: The effects of cohesion and range [Publisher: SAGE Publications]. *Administrative science quarterly*, *48*(2), 240–267.

Rogers, E. M. (2003). *Diffusion of Innovations, 5th Edition* (5th edition). Free Press.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, *115*(3), 211–252. https://doi.org/10.1007/s11263-0150816-y

Ryan, B., & Gross, N. (1950). Acceptance and diffusion of hybrid corn seed in two Iowa communities. *Research Bulletin (Iowa Agriculture and Home Economics Experiment Station)*, *29*(372), 49.

Samadi, S., Tantipongpipat, U., Morgenstern, J., Singh, M., & Vempala, S. (2018). The Price of Fair PCA: One Extra Dimension [arXiv:1811.00103 [cs, stat]]. https://doi.org/10.48550/arXiv.1811.00103 Saraswathi, S., Mukhopadhyay, A., Shah, H., & Ranganath, T. S. (2020). Social Network Analysis of COVID19 Transmission in Karnataka, India. *medRxiv*, 2020.08.11.20172734. https://doi.org/10.1101/2020. 08.11.20172734

Schnabel, T., Swaminathan, A., Singh, A., Chandak, N., & Joachims, T. (2016). Recommendations as Treatments: Debiasing Learning and Evaluation [arXiv:1602.05352 [cs]]. https://doi.org/10.48550/arXiv. 1602.05352

Schneider, C. M., Mihaljev, T., Havlin, S., & Herrmann, H. J. (2011). Suppressing epidemics with a limited amount of immunization units. *Phys. Rev. E*, *84*(6), 061911. https://doi.org/10.1103/PhysRevE. 84.061911

Small, M. L. (2004). *Villa Victoria: The Transformation of Social Capital in a Boston Barrio* (1st edition). University of Chicago Press.

Small, M. L. (2010). *Unanticipated Gains: Origins of Network Inequality in Everyday Life* (Reprint edition). Oxford University Press.

Sorenson, O., Rivkin, J. W., & Fleming, L. (2006). Complexity, networks and knowledge flow. *Research Policy*, *35*(7), 994–1017. https://doi.org/10.1016/j.respol.2006.05.002

Stoica, A.-A. (2020). Algorithmic Fairness for Networked Algorithms. *AAMAS*, 2214–2216.

Stoica, A.-A., & Chaintreau, A. (2019). Fairness in Social Influence Maximization. In S. Amer-Yahia, M. Mahdian, A. Goel, G.-J. Houben, K. Lerman, J. J. McAuley, R. Baeza-Yates, & L. Zia (Eds.), *Companion of The 2019 World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019* (pp. 569–574). ACM. https://doi.org/10.1145/3308560.3317588

Stoica, A.-A., Han, J. X., & Chaintreau, A. (2020). Seeding Network Influence in Biased Networks and the Benefits of Diversity. In Y. Huang, I. King, T.-Y. Liu, & M. v. Steen (Eds.), *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020* (pp. 2089–2098). ACM / IW3C2. https://doi.org/10.1145/3366423.3380275

Tsang, A., Wilder, B., Rice, E., Tambe, M., & Zick, Y. (2019). Group-Fairness in Influence Maximization. In S. Kraus (Ed.), *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019* (pp. 5997–6005). ijcai.org. https://doi.org/10.24963/ijcai.2019/831

Valente, T. W., & Rogers, E. M. (1995). The origins and development of the diffusion of innovations paradigm as an example of scientific growth. *Science Communication*, *16*(3), 242–273. https://doi.org/10. 1177/1075547095016003002

Valente, T. W. (1996). Social network thresholds in the diffusion of innovations. *Social Networks*, *18*(1), 69–89. https://doi.org/10.1016/0378-8733(95)00256-1

Valente, T. W. (2012). Network interventions. *Science (New York, N.Y.)*, *337*(6090), 49–53. https://doi.org/10.1126/science.1217330

Valente, T. W., & Davis, R. L. (1999). Accelerating the diffusion of innovations using opinion leaders [Publisher: Sage Publications]. *The Annals of the American Academy of Political and Social Science*, *566*(1), 55–67.

Valente, T. W., & Fujimoto, K. (2010). Bridging: Locating Critical Connectors in a Network. *Social networks*, *32*(3), 212–220. https://doi.org/10.1016/j.socnet.2010.03.003

Valente, T. W., Hoffman, B. R., Ritt-Olson, A., Lichtman, K., & Johnson, C. A. (2003). Effects of a SocialNetwork Method for Group Assignment Strategies on Peer-Led Tobacco Prevention Programs in Schools. *American Journal of Public Health*, *93*(11), 1837–1843. Retrieved September 5, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1448060/

Valente, T. W., & Pitts, S. R. (2017). An Appraisal of Social Network Theory and Analysis as Applied to Public Health: Challenges and Opportunities. *Annual Review of Public Health*, *38*, 103–118. https://doi.org/10.1146/annurev-publhealth-031816-044528

Valente, T. W., & Pumpuang, P. (2007). Identifying opinion leaders to promote behavior change. *Health*

Education & Behavior: The Official Publication of the Society for Public Health Education, *34*(6), 881–896. https://doi.org/10.1177/1090198106297855

Valente, T. W., Ritt-Olson, A., Stacy, A., Unger, J. B., Okamoto, J., & Sussman, S. (2007). Peer acceleration: Effects of a social network tailored substance abuse prevention program among high-risk adolescents [eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1360-0443.2007.01992.x]. *Addiction*, *102*(11), 1804–1815. https://doi.org/10.1111/j.1360-0443.2007.01992.x

Valente, T. W., & Vega Yon, G. G. (2020). Diffusion/Contagion Processes on Social Networks [Publisher: SAGE Publications Inc]. *Health Education & Behavior*, *47*(2), 235–248. https://doi.org/10.1177/1090198120901497

Van den Bulte, C., & Joshi, Y. V. (2007). New Product Diffusion with Influentials and Imitators [Publisher: INFORMS]. *Marketing Science*, *26*(3), 400–421. Retrieved September 6, 2023, from https://www.jstor.org/stable/40057065

van de Rijt, A., Siegel, D., & Macy, M. (2009). Neighborhood Chance and Neighborhood Change: A Comment on Bruch and Mare [Publisher: The University of Chicago Press]. *American Journal of Sociology*, *114*(4), 1166–1180. https://doi.org/10.1086/588795

Venkatasubramanian, S., Scheidegger, C., Friedler, S., & Clauset, A. (2021). Fairness in Networks: Social Capital, Information Access, and Interventions. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 4078–4079. https://doi.org/10.1145/3447548.3470821

Verma, S., & Rubin, J. (2018). Fairness definitions explained. In Y. Brun, B. Johnson, & A. Meliou (Eds.), *Proceedings of the International Workshop on Software Fairness, FairWare@ICSE 2018, Gothenburg, Sweden, May 29, 2018* (pp. 1–7). ACM. https://doi.org/10.1145/3194770.3194776

Wang, W., & Street, W. N. (2018). Modeling and maximizing influence diffusion in social networks for viral marketing [Number: 1 Publisher: SpringerOpen]. *Applied Network Science*, *3*(1), 1–26. https://doi.org/10.1007/s41109-018-0062-7

Watts, D. J. (1999). Networks, Dynamics, and the Small-World Phenomenon. *American Journal of Sociology*, *105*(2), 493–527. https://doi.org/10.1086/210318

Watts, D. J., & Dodds, P. S. (2007). Influentials, networks, and public opinion formation [Publisher: The Oxford University Press]. *Journal of consumer research*, *34*(4), 441–458.

Western, B., & Muller, C. (2013). Mass Incarceration, Macrosociology, and the Poor [Publisher: SAGE Publications Inc]. *The ANNALS of the American Academy of Political and Social Science*, *647*(1), 166–189. https://doi.org/10.1177/0002716213475421

Wyly, E. K., Atia, M., Foxcroft, H., Hamme, D. J., & Phillips-Watts, K. (2006). American home: Predatory mortgage capital and neighbourhood spaces of race and class exploitation in the United States. *Geografiska Annaler: Series B, Human Geography*, *88*(1), 105–132. https://doi.org/10.1111/j.04353684.2006.00208.x

Yadav, A., Wilder, B., Rice, E., Petering, R., Craddock, J., Yoshioka-Maxwell, A., Hemler, M., Onasch-Vera, L., Tambe, M., & Woo, D. (2018). Bridging the Gap Between Theory and Practice in Influence Maximization: Raising Awareness about HIV among Homeless Youth, 5399–5403. Retrieved July 2, 2021, from https://www.ijcai.org/proceedings/2018/761

Yang, T., & Aldrich, H. E. (2014). Who's the Boss? Explaining Gender Inequality in Entrepreneurial Teams [Publisher: SAGE Publications Inc]. *American Sociological Review*, *79*(2), 303–327. https://doi.org/10.1177/0003122414524207

Young, L. E., Sidnam-Mauch, E., Twyman, M., Wang, L., Xu, J. J., Sargent, M., Valente, T. W., Ferrara, E., Fulk, J., & Monge, P. (2021). Disrupting the COVID-19 Misinfodemic With Network Interventions: Network Solutions for Network Problems. *American Journal of Public Health*, *111*(3), 514–519. https://doi.org/10.2105/AJPH.2020.306063

Zafar, M. B., Valera, I., Gomez-Rodriguez, M., & Gummadi, K. P. (2017). Fairness Beyond Disparate Treatment & Disparate Impact: Learning Classification without Disparate Mistreatment. *WWW 2017*, 1171–1180.

Zhang, J., Brackbill, D., Yang, S., Becker, J., Herbert, N., & Centola, D. (2016). Support or competition? How online social networks increase physical activity: A randomized controlled trial. *Prev Med Rep*, *4*, 453–458. https://doi.org/10.1016/j.pmedr.2016.08.008

Zhang, J., Brackbill, D., Yang, S., & Centola, D. (2015a). Efficacy and causal mechanism of an online social media intervention to increase physical activity: Results of a randomized controlled trial. *Prev Med Rep*, *2*, 651–657. https://doi.org/10.1016/j.pmedr.2015.08.005

Zhang, J., Brackbill, D., Yang, S., & Centola, D. (2015b). Identifying the effects of social media on health behavior: Data from a large-scale online experiment. *Data in Brief*, *5*, 453–457. https://doi.org/10.1016/j.dib.2015.09.049

Zhang, J., & Centola, D. (2019). Social Networks and Health: New Developments in Diffusion, Online and Offline. *Annu. Rev. Sociol.*, *45*(1), 91–109. https://doi.org/10.1146/annurev-soc-073117-041421

Zuckerman, E. W., & Jost, J. T. (2001). What Makes You Think You're so Popular? Self-Evaluation Maintenance and the Subjective Side of the "Friendship Paradox" [Publisher: [Sage Publications, Inc., American Sociological Association]]. *Social Psychology Quarterly*, *64*(3), 207–223. https://doi.org/10.2307/3090112

Zweigenhaft, M. R. L., & Domhoff, P. G. W. (1999). *Diversity in the Power Elite: Have Women and Minorities Reached the Top?* Yale University Press.