Norms Without Knowledge: Intervention Heuristics for Influencing Social Norms in Networks

William Brown

Submitted to the Department of Philosophy at the University of Pennsylvania in partial fulfillment of the requirements for Honors.

Thesis Advisor: Professor Cristina Bicchieri

Department of Philosophy
University of Pennsylvania
We consider the question of how to optimally influence the adaptation of positive social norms in a community, given limited information about the community structure and a limited budget for intervention. We focus primarily on a game-theoretic model of normative behavior in a social network which permits an analytic solution for both the equilibrium action profile and the intervention which maximizes social welfare. Computing the optimal intervention requires full knowledge of the adjacency matrix for the network, which may be unrealistic in practical scenarios. We propose several heuristic approaches for intervention which can be implemented with limited information about the network structure and conduct simulations of these heuristics to evaluate their efficacy. Our results illustrate the strengths and tradeoffs of these heuristics under various network conditions and can serve guidelines for promoting positive norms in communities.
Introduction

Humans are social animals. To a large extent, the choices we make and actions we take are driven by beliefs about the expectations of others in our community. This often results in the propagation of behaviors which bring detrimental effects to those who practice them. Consider, for example, a social norm of truancy in a high school friend group. If one finds that a large number of their closest friends routinely skip a particular class in favor of more immediately gratifying adventures, the pressure to accompany them can be immense. This pressure consists not only of a belief that skipping class would be enjoyable; the truants may all be well aware that they are harming themselves in the long run by missing out on pieces of their education or harming their grades. Rather, once the behavior is established, each individual may feel an expectation from their friends that they continue their haphazard attendance, and that they would face social ostracization and loss of friendships if they reject the social norm of truancy. It would be a mistake to characterize these choices as simply irrational lapses of judgment on behalf of each individual. If those friendships are so important that they constitute a large portion of one’s joyful experiences, social or otherwise, skipping class may be the rational or utility-optimizing action for these individuals, conditioned on their expectations of repercussions upon deviating.

Now suppose that this norm of truancy is becoming prevalent in numerous friend groups throughout a particular high school, culminating in decreased test performance and graduation rates. What should the principal of this school do to mitigate this behavior and raise attendance rates? Broadly deployed systems of reward and punishment may have some effect, but they do nothing to address the underlying social factors which spur the behavior in the first place. The principal may have luck in shifting the short-term behavior of a particular individual if they sat down for a conversation with this person, discussed their attendance record and motivations for truancy, reminded them of the importance of education for future success, and developed a personal relationship which provided an additional incentive to come to class. Yet it would be impossible to have this interaction with every student in the school. And if this student begins attending class more frequently while their friends remain absent, and their friendships suffer as a result, they may soon revert to their original behavior. However, if a critical mass of students increased their attendance following direct interventions, those in their immediate friend groups may follow suit, thereby degrading individuals’ beliefs that others expect them to skip class. To eradicate the norm of truancy, a goal of the principal could be to identify a small group of students for whom an intervention and subsequent behavioral change would prompt a cascade of defection from the norm throughout the school.
Social norms need not exclusively encourage harmful behaviors. A community may exhibit a social norm for engaging in political activism, volunteer work, or giving charitable donations to worthwhile causes. Engaging in each of these actions requires an investment on behalf of the individual in terms of time or money, but the resulting effects can be tremendously positive for the community if practiced at scale. A social norm for such behavior can be crucial for its maintenance; indeed, many communities exhibit a strong social stigma against not voting or not donating a certain portion of income to charity. If a non-profit organization wants to instill such a norm in a community which is somewhat apathetic at present, they are tasked with a similar objective as the aforementioned high school principal. Individuals will be much more likely to engage in the desired behaviors if they perceive social pressure to do so, and so the organization must identify an optimal strategy for spurring change in the community, subject to a constraint on resources for intervention.

For this thesis, we place ourselves in the shoes of a benevolent third party who wishes to promote positive change in a community. We review several approaches for modeling community behavior and focus primarily on the model from Galeotti et al. [7]. In this model, individuals in a community are represented as nodes in a network, where edges represent connections between individuals, possibly with weights dictating relative strengths of connections. Each agent chooses the action which maximizes their utility, which is given in terms of their individual reward for the action, the cost of taking the action, and the social reward or cost for the action, which is a function of the actions of their neighbors. Interventions are captured in this model by assuming that the parameters dictating the individual reward term in each agent’s utility function can be modified, with a budget constraint on the amount of reward modification throughout the network. The intervention objective is simply to maximize social welfare, or the sum of all individual utilities. In the Galeotti et al. model, it is possible to compute both the equilibrium action profile for all players as well as the optimal intervention for a budget when the structure of the network is fully known. However, acquiring this information may be expensive or impossible in practice. How could a high school principal or non-profit organization proceed in mapping the entire social structure of the community in which they wish to intervene? If this is information is unobtainable, is all hope lost in designing a successful and cost-effective intervention strategy? Addressing the latter question is our central aim. Rather than attempting to discover the single best intervention, we seek heuristic approaches which perform reasonably well (under a variety of network conditions, possibly making use of randomization). We give several heuristic strategies, drawing on heuristic approaches for influencing other kinds of network behavior, and run simulations of their implementation in the Galeotti et al. model. Our results show that the barriers of lacking network data can be largely overcome with simple heuristics, and
that we can match the performance of the optimal intervention with a very small increase in our budget.

**Background: Game Theory and Networks**

Before delving into the literature on social norms, we begin by discussing a few important topics in game theory and networks. These concepts will prove useful in providing a mathematical characterization for qualitative theories about social norms.

**Equilibria and Best Response Dynamics**

Central to the study of games is the notion of a Nash equilibrium point. Here, a “game” simply refers to a set of agents and actions, with a function specifying the utility of each agent in terms of the actions chosen by all agents. An action profile specifying the single action (pure strategy) or random distribution over actions (mixed strategy) taken by each player is a Nash equilibrium if no player can improve their utility by changing their action, conditioned on the actions of all other players remaining the same. In the classic game of Rock-Paper-Scissors (where winning yields utility of 1, losing yields -1, and a tie results in utility of 0 for both players), the single Nash equilibrium action profile is to play all actions randomly each with probability 1/3; if you play any action with a higher probability (say, Scissors at 50%), your opponent can win at a higher rate by playing the corresponding dominating action constantly (Rock at 100%). John Nash’s landmark PhD thesis showed the existence of equilibria in all two player games with a finite number of actions [10]; a wide variety of other games are guaranteed to have equilibrium points as well. When complex real world scenarios can be modeled as games, studying equilibrium points can give insight into the actions which are “stable” in that scenario.

Yet simply because an equilibrium exists in a game, why should we expect that players arrive at this equilibrium through some “natural” progression of play? It turns out that in a broad class of games where pure strategy equilibria exist, players will arrive at such a profile through a simple iterative procedure known as Best Response Dynamics [6]. Suppose that a game is not presently at equilibrium; this means that at least one player can improve their utility by changing their action in isolation. In Best Response Dynamics, this update simply repeats until no player can improve their outcome by deviating. This might align with our natural intuitions about the behavior of agents in competitive scenarios, where each agent makes myopic improvements to their strategy whenever they see an opportunity. One way to show the convergence of Best Response Dynamics to a Nash equilibrium is by showing
that there is an “ordinal potential function”, or a function of the players’ actions which only increases whenever a player makes a best response. It is not too difficult to see why this results in convergence. Any function of the game state must have some maximum value, and thus cannot increase indefinitely. As players repeatedly deviate to improve their utility, the potential function increases towards its maximum value. At this value, because no player can improve their utility (doing so would imply that the current potential function value is not the maximum value), the resulting action profile will be a Nash equilibrium. In some games, the sum of all agents’ utilities suffices for an ordinal potential function. We will later see that this kind of argument can be used to justify the efficacy of the interventions we consider.

**Braess’ Paradox: An Illustrative Example**

While myopic utility-optimizing actions by each agent can lead to a Nash equilibrium, this does not necessarily mean that the resulting equilibrium leads to a good outcome in terms of social welfare. The existence of suboptimal Nash equilibria is evident from the canonical Prisoners’ Dilemma. We illustrate this further with an example which more closely resembles a scenario regarding a large community of people interacting with each other, albeit without an explicit representation of social norms. In Braess’ paradox [4], a routing problem in a network is shown to have a lower equilibrium social welfare when edges are added to the network. We can think of this as a scenario where agents are choosing a driving route to reach some destination from the same starting point; here, Braess’ paradox implies that building additional roads can result in worse traffic. In the example, we let the node $S$ be the starting point for a population of 100 agents, all of whom want to arrive at node $T$ in the shortest time possible. The values on each edge dictate the time required for traversal, where $x$ denotes the number of agents who choose to travel along that edge. We can think of the edges $(S, A)$ and $(B, T)$ as routes through a city which are quite fast in low traffic, but become much slower under congestion. The edges $(S, B)$ and $(A, T)$ are more analogous to highways, which can still allow drivers to travel at their maximum speeds even under high traffic.

\[
\begin{array}{c}
S \\
\downarrow \frac{x}{100} \\
A \\
\downarrow 1 \\
B \\
\downarrow 1 \\
T \\
\uparrow \frac{x}{100}
\end{array}
\]

Here, the equilibrium behavior would be for half of the agents to pass through each of $A$
and $B$, all achieving a travel time of 1.5, thus resulting in a total travel time of 150. If an agent passing through $B$ decided to deviate to the route through $A$, their travel time would increase to 1.51, and so they should choose to maintain their current strategy; the same holds for agents currently passing through $B$. Now suppose that a city planner decides to build an underground tunnel between $A$ and $B$ which allows instantaneous transit. How will this affect the equilibrium behavior of the drivers?

![Graph](attachment:image.png)

Under this new network design, the resulting equilibrium is actually worse for all agents, despite the fact that the capacity of the network has only been increased. To see this, observe that for any agent originally passing through $B$, it will never increase their travel time to move through $A$ first and subsequently traverse the new $(A, B)$ edge, as $\frac{x}{100} + 0 \leq 1$ for all values of $x$ below 100. Likewise, any driver already passing through $A$ can only improve their travel time if they choose to additionally pass through $B$ instead of traveling directly to $T$. In the action profile where all drivers pass through the $(A, B)$ edge, each agent obtains a travel time of 2. As no one can improve their travel time through deviating, this results in a Nash equilibrium with total travel time of 200, which is considerably worse than before the new tunnel was built. Best Response Dynamics would converge to this point if agents began at the old equilibrium; the first agent to deviate to following the $(A, B)$ edge would initially reduce their travel time to 1.01, but this would slowly rise to the equilibrium value of 2 as all other agents deviated to reduce their own travel time.

Not only are scenarios like this important fact to keep in mind when designing road plans, but they can serve as an analogy for the problem of intervening in a community to improve social welfare. An outside authority seeking to improve commuting times in this updated traffic network could simply decide to close off the tunnel and allow agents to return to their old, faster driving patterns. The idea that social welfare can be far from optimal is described vividly by the notion of the “Price of Anarchy” for a game [5]. The Price of Anarchy is simply the ratio between the optimal achievable social welfare at any point, which need not be an equilibrium, and the social welfare at the worst possible equilibrium. This captures the idea that systems operating without direct intervention can result in much worse outcomes for the involved agents compared to a scenario where a benevolent dictator could prescribe all actions. Yet it may be impossible to simply dictate to all agents the action that they must
take. Rather, the form of intervention we study can be viewed as *changing the rules of the game* such that the resulting equilibrium improves overall social welfare. This is effectively what is happening if we remove the \((A, B)\) edge in the example above.

**PageRank and Network Centrality**

In the behavioral model we use for our experimentation, interventions will take the form of an adjustment to the utility function of particular individuals in the network. The model assumes that we can expend effort to shift the utility function of an individual in such a way that they are more likely to take the desired action. In the example of a non-profit trying to promote a culture of activism in a community, this might look like using targeted advertisements to convince people that volunteering would be a worthwhile and fulfilling experience. However, ads can become expensive when deployed at scale, and we might hope to do better than blindly selecting people at random. How can we determine the most important people in a network, who would be likely become effective influencers and community leaders upon adopting the target behavior? These won’t necessarily be the people with the most immediate connections. Someone who is tightly connected to a small group on the fringe of a network may have more neighbors than someone who is loosely connected throughout the network, but would still be less influential on the whole. Further degrees of separation are important factors in a person’s influence, and so we need to be able to consider the entire structure of the network in determining those who we deem important.

Larry Page and Sergey Brin, the founders of Google, found themselves facing a similar problem while attending graduate school at Stanford [CITE - pop sci article]. At the time, many search engines operated simply by attempting to find websites with a large number of words in common with an entered query. These search engines were relatively easy to reverse engineer; a website could artificially embed words from common queries into their page to boost their rankings on the results page. Page and Brin sought to leverage the network structure of the World Wide Web to determine important web pages [CITE - PR paper]. Here, the edges in the network are represented by embedded links from one website to another. If a website is referenced by numerous other sites that are known to be important, this indicates that this website is relatively important as well. But how can we determine which websites are important in the first place? Page and Brin devised the PageRank algorithm for accomplishing precisely this goal, which eventually ended up serving as the primary ranking algorithm for Google search results. In PageRank, every page is initially allocated the same amount of “influence”. In each iteration, all pages disperse their influence evenly across their outgoing links to other pages, in turn receiving influence from incoming
Quantitative Models of Social Norms

While discussions on the importance of norms and social influence date back to Plato and Aristotle, the practice of developing quantitative models for social norms began in earnest in the latter half of the twentieth century. Quantitative modeling requires a precise understanding of that which one hopes to model; in *Norms in the Wild*, Bicchieri [2] defines a social norm as a “rule of behavior such that individuals conform to it on condition that they believe that (a) most people in their reference network conform to it (empirical expectation), and (b) that most people in their reference network believe they ought to conform to it (normative expectation)” (35). We briefly review several important models of social norms which share semantic similarities to this definition.

The Return Potential Model from Jackson (1965) [9] describes a curve in two dimensions, characterizing the expected community response to an action which can be partaken in varying degrees; for example, a driver may experience a negative reaction from law enforcement or other drivers for driving substantially above or below the listed speed limit for a road. Selecting an action which maximizes the value of the curve would correspond to following the corresponding social norm.

Schelling’s model of segregation (1971) [5] was designed to provide an explanation for the prevalence of housing segregation by race and to illustrate the difficulties which must be overcome to combat it. Segregation can be viewed as a social norm, where individuals may choose to self-segregate even if they have no personal preference for living among individuals of the same race. The presence of normative expectations can be strong enough to induce segregation in a population where a large number of people do not personally support the norm, and the Schelling model enables simulations for resulting geographic distributions of race according to various parameter values describing the strength of normative expectations.

Granovetter’s landmark paper “Threshold Models of Collective Behavior” (1978) [8] provides a model for understanding cascades of behavior adoption in a population. For any behavior for which a social norm exists, individuals may have a preference to partake in the behavior which is conditional on the proportion of their peers who also partake. If there are 100 people in an area where a riot is forming, and I am willing to join the riot only if 50 or
more people have joined, then my threshold for participation in this model is 50%. If we can estimate the distribution of behavior thresholds within a population as well as the number participating in the behavior at a given time, we can use this model to estimate whether the behavior is likely to explode or disperse. By delineating discrete time intervals, we can inductively calculate estimates of the proportion that participate at each interval, which will eventually converge to a stable equilibrium.

All of these models abstract away many of the details from real world scenarios where social norms are present; true human behavior is noisy, influenced by emotion, and not fully describable by a simple equation. Yet, for understanding the broad effects of social norms, these models are quite powerful. With appropriate parameter values, the predictions of these models are certainly correlated with the true behavior they aim to capture, and can provide useful signals about how we might hope to influence normative behavior. We will further discuss two quantitative models for social norms which are directly relevant to the task of influencing behavioral adoption in structured social networks.

The Bicchieri Model

In *Norms in the Wild*, Bicchieri [2] discusses the importance of trendsetters for shifting social norms in networks. She argues that trendsetters are typically agents with “low risk sensitivity, low risk perception, low allegiance to the standing norm, high autonomy, and high perceived self-efficacy” (163), and that these agents typically are found on the periphery of a network. In “Norm Change: Trendsetters and Social Structure” [3], Bicchieri and Funcke develop a quantitative model of social norms which showcases the importance of trendsetters. Here, agents in a network can choose between two actions, either to Follow or Transgress from the norm ($F$ and $T$). A payoff function $\pi : k_i \times \sigma_1 \times \sigma_2 \rightarrow U$ determines the utility of an agent in a pairwise game with each of its neighbors depending on their actions. The parameter $k_i$ controls the agent’s sensitivity to the prevalent norm; agents with high sensitivity experience larger decreases in utility when they do not coordinate with others. The parameter $x$ represents an agent’s belief about the prevalence of a norm in the relevant subpopulation of a network. Given $x$, an agent’s expected utility for each action is:

$$E[U(F)|x] = x\pi(k_i, F, F) + (1 - x)\pi(k_i, F, T)$$
$$E[U(T)|x] = (1 - x)\pi(k_i, T, T) + x\pi(k_i, T, F)$$

The parameter $\beta$ captures the relative importance of the local and global network for agents’ beliefs. This is used in each agent’s computation of $x$. Agents consider their expectations
of future outcomes in their action choices, aiming to maximize expected long-run utility. Bicchieri and Funcke run simulations of this model in a social network, where the prevalence of each behavior is computed over many time steps. Their results validate their hypotheses about which agents are likely to be trendsetters, and they indentify conditions under which the trendsetters are likely to be successful in spreading their behavior throughout the network.

The Galeotti et al. Model

Andrea Galeotti, Benjamin Golub, and Sanjeev Goyal [7] provide another model for norm dynamics in networks which captures the social aspects of behavioral adoption that we aim to study. Their model shares many similarities with the Bicchieri model, and has additional properties which are useful for studying interventions. Namely, their model permits analytic computation of the Nash equilibrium action profile for all agents, as well as the optimal intervention subject to a budget constraint. We describe the details of the model below.

Setting

In this setting, agents are playing a game on a network with adjacency matrix $G$, where $g_{ij}$ is a value representing the strength the connection between $i$ and $j$. Each agent takes an action $a_i \in \mathbb{R}$ and solely aims to maximize their individual welfare. We think of these actions which can be followed in degrees, such as time or energy spent on some activity. The welfare for an agent $i$ is given by:

$$w_i = b_i a_i - \frac{1}{2} a_i^2 + \beta \sum_j g_{ij} a_i a_j$$

Each of the three terms in the welfare equation corresponds to a different influencing factor in an individual’s choice of action; we interpret them as loosely corresponding to the individual reward received, the effort exerted, and the social cost or reward for taking an action. In the first term, the value $b_i$ is a unique multiplier for each agent which represents the agent’s private value for increasing their level of action. The second term reflects the cost of increasing the level of action. The quadratic growth of this term can be interpreted as indicating that the action exhibits increasing marginal cost at higher degrees; absent any influence from the network via the third term, an isolated agent would maximize welfare by choosing $a_i = b_i$, which is the sole value where $\nabla w_i = 0$. The third term reflects the normative aspect of the model. The parameter $\beta$ controls the network dynamics, where a positive
sign promotes cooperative behavior with your neighbors and a negative sign promotes acting opposite your neighbors. We focus solely on cooperative behavior, which aligns with the notion of normative expectations. Agents reap additional rewards for high levels of action if such levels are also exhibited in their reference network, and thus each agent’s preference for a high level of action is conditional on the prevalence of the action. This model makes the true actions of one’s neighbors public to each agent; this creates an implicit assumption that each agent’s empirical and normative expectations are accurate. As such, our goal is not to distort the beliefs of agents in hopes that they adopt a norm, but rather to shift their individual reward function in favor of preference for the norm.

Computing Nash Equilibria

The Nash equilibrium action profile in the Galeotti et al. model can be computed directly via the specified parameters for the agent preferences and network structure. Differentiating with respect to the action vector, the Best Response vector can be expressed as:

$$\text{BR} = \mathbf{b} + \beta \mathbf{G} \mathbf{a}$$

Using the modified adjacency matrix $M = (I - \beta \mathbf{G})$, where $I$ is the identity matrix, upon solving for $\text{BR} = \mathbf{a}$ we get:

$$\mathbf{a} = (I - \beta \mathbf{G})^{-1} \mathbf{b} = M^{-1} \mathbf{b}$$

As all agents are best-responding to each other, and the computed vector is unique, this is the sole pure strategy Nash equilibrium for the network. We can also see that Best Response Dynamics results in convergence to this equilibrium by observing that the sum of all agents’ welfare is a potential function. Suppose $\beta$ and all $\mathbf{b}_i$ values are positive, and all agents start with initial action levels $a_i = 0$. This is below the optimal action level for all agents, and each agent’s utility function is concave upon fixing all other action values. When each agent best responds, they will increase their action level towards the optimum value. Because $\beta$ is positive, these increases can only improve the welfare of all other agents. Social welfare increases upon every best response, and thus best response dynamics will converge.

Finding the Optimal Intervention

The mechanism for intervention in the Galeotti et al. model supposes a central authority has the ability to modify agents’ private multipliers for an action; to shift an agent’s multiplier from $\mathbf{b}_i$ to $\hat{\mathbf{b}}_i$, they can pay $(\mathbf{b}_i - \hat{\mathbf{b}}_i)^2$, and the total cost paid must be bound by $C$. Similar to the phenomenon of increasing marginal cost for agents increasing their levels of action,
this quadratic growth also assumes increasing marginal cost in efforts to shift an agent’s utility function. The goal of the central authority is to maximize social welfare subject to the budget constraint, i.e.: 

$$\max \sum_i w_i \text{ subject to } \sum_i (b_i - \bar{b}_i)^2 \leq C$$

Their work presents a method for computing the optimal budget-constrained in terms of the eigenvectors of $G$. We omit the full technical details of this computation, but we note that for large values of $C$, the proportions of the intervention are dominated by the top eigenvector of $G$. This eigenvector is an important measure of network centrality, and is closely related to the PageRank vector for a network. Interestingly, this optimal intervention is not dependent on the private weights of agents; from the formulation of the best response computation above, we can see that the shift in an agent’s optimal action resulting from a shift in their multiplier $b_i$ will be the same regardless of their initial value for $b_i$. When $\beta > 0$, we can again see that Best Response Dynamics would result in convergence to the new Nash equilibrium after updating the $b_i$ values. All updates will be positive, and thus the social welfare of the network remains a potential function for the game.

While it would be ideal to be able to determine the optimal intervention strategy for influencing social norms, in practice we do not have access to the full adjacency matrix for a network of interest. Our experimental simulations focus on attempting to overcome this information barrier and achieve results via heuristic approaches which are close to optimal. The optimal intervention will typically also involve some shift in multipliers for all agents in the network. Again, this may be infeasible in practice, particularly if we lack identifying information for all agents. Our simulations explore various heuristics for targeted interventions with constraints on the number of agents we are allowed to make contact with, in addition to the standard budget constraint.

**Similarities to the Bicchieri Model**

The cooperative action case ($\beta > 0$) bears many similarities to the Bicchieri model, as well as some crucial differences. The different values in $G$ can be used to represent strength of influence from different members in the network. Just as the $\beta$ parameter in the Bicchieri model controls the depth in a network which agents look to form their beliefs, we can use values of $G$ which represent the phenomenon of being influenced by those other than your direct neighbors. The $b_i$ values play a similar role to the $k_i$ values in the Bicchieri model, controlling an agent’s relative sensitivity to social norms. The models differ in their
representation of trendsetters and their predictions about the likelihood of agents to be influential in shifting norms. A trendsetter in the Bicchieri model would be an agent with low connectivity and a high $b_i$ value in the Galeotti et al. model. The Galeotti et al. model would not select these agents as being the most influential, in part because they ignore the $b_i$ values in targeting interventions, but primarily because these agents would not have high network centrality.

This does not necessarily mean that the models of incompatible. The Bicchieri model primarily aims to describe the beliefs of agents about others’ normative expectations to predict which agents are likely to deviate. The Galeotti et al. model abstracts away the mechanics of these belief systems by assuming agents can directly observe the behaviors of their neighbors and assuming that agents play a Nash equilibrium. The aforementioned Best Response Dynamics protocol for arriving at equilibrium the Galeotti et al. could be seen as the analog of the rounds of updating in the Bicchieri model. The goal of the Galeotti et al. model is more prescriptive. Rather than trying to predict where change will emerge from, the aim of Galeotti et al. is prescriptive, in that they assume a third party outside the network will be the ultimate source of change. It could still be the case that the agents more likely to exhibit the normative behavior before an intervention is applied in the Galeotti et al. model are those who would be trendsetters in the Bicchieri model. But that does not necessarily imply that additional motivational effort would be best spent on the trendsetters rather than more central agents.

The Value of Network Data

Recent work by Akbarpour, Malladi, and Saberi [1] has also focused on the challenge of overcoming network data barriers in targeted interventions. They study the related problem of behavioral seeding in social networks, but their techniques focus on settings where norms are descriptive rather than social. Here, the probability that an agent adapts a norm is simply a function of the prevalence of a norm in their reference network. As such, their model behaves much like a contagion model for disease or information, where contact with others is the primary indicator of whether the norm spreads. They presume that the aim of a third party is to disseminate some norm in a network where the norm is nonexistent, and that they can “seed” the norm with a small number of individuals, who adopt the behavior as a result. The challenge here is to determine the optimal agents to seed in order to maximize dispersal. Under their model, the optimal agents can be computed directly from the adjacency matrix for the network; however, their primary aim is to explore the efficacy of randomized strategies when the edges of the network are unknown. In this setting, for
Erdos-Renyi random graphs, they derive theoretical bounds on the additional effort needed to match optimal targeting when network information is limited. We experimentally show qualitatively similar results for the problem of influencing social norms in the Galeotti et al. model.

**Experimental Design**

We construct a Python implementation of the Galeotti et al. model which enables the creation of random networks, parameter assignment, computation of Nash equilibria, and computation of social welfare. Our implementation additionally allows us to evaluate the impacts to social for any intervention as well as compute the dominant component of the optimal intervention. In addition to the optimal intervention, we propose five heuristic approaches, requiring varying degrees of information about the network structure.

**Proposed Heuristics**

We let $N$ denote the number of agents in the network. Most of these heuristic methods are parameterized by a value $K$, which indicates the number of agents whose utility functions are modified by the intervention. Presented below are descriptions for each of the interventions:

- **Uniform**: The $b_i$ values for all agents are increased by $\sqrt{C/N}$, as this results in a total intervention cost of $C$.

- **Random-$K$**: The $b_i$ values for $K$ agents, selected randomly without replacement, are increased by $\sqrt{C/K}$.

- **Friends-$K$**: In this heuristic, we aim to select agents with high network centrality without knowledge of the edges in the graph. We choose $K$ agents by selecting an agent at random, including a random neighbor of that agent in our set, and repeating until $K$ distinct agents have been selected. This biases our sample towards agents with high degree, as they are more likely to be the neighbor of a randomly selected vertex. The $b_i$ values for these $K$ agents are increased by $\sqrt{C/K}$.

- **Top-$K$**: The $b_i$ values for the $K$ agents with the most neighbors are increased by $\sqrt{C/K}$.

- **PageRank-$K$**: The $b_i$ values for the $K$ agents with the largest PageRank values are increased by $\sqrt{C/K}$.
For comparison, we also consider the optimal intervention as well as a baseline where no intervention is applied:

- **OPT:** The $b_i$ values for all agents are increased in proportion with the dominant eigenvectors of $G$, as described in [7].

- **Baseline:** All $b_i$ values are unchanged.

We simulate the applications of these interventions in Erdos-Renyi graphs with varying values for $C$ and $K$, as well as varied distributions of the $b_i$ values for each agent.

**Results**

For all experiments and for specification of parameter values for $C$ and $K$, we run multiple trials, randomizing over the construction of the network, the selection of targeted agents by our heuristics, and the distributions for agents’ $b_i$ values. We fix the size of the network at 100 and include every edge with a 10% probability. The value for $\beta$ is fixed at 0.1, reflecting a moderate level of normative expectation for conforming on behalf of agents with high levels of action. Edges between agents $i$ and $j$ are represented by setting $g_{ij} = 1$; all values in the adjacency matrix corresponding to non-existent edges are set to 0.

**Varying $K$**

We first demonstrate the efficacy of our intervention heuristics as we vary the value of $K$. Our outcome variable of interest is the total network utility, or $\sum_i w_i$. We repeat this experiment with three different distributions for agents’ $b_i$ values. For the results shown in Figure 1, all $b_i$ values are set to 1, reflecting uniform preferences for adopting the norm behavior in isolation. For the results shown in Figure 2, these values are drawn from a uniform distribution over $[0, 2]$. For Figure 3, the values are drawn from a normal distribution with mean 2 and standard deviation 0.5; these values are chosen to ensure that with high likelihood, all agents will have positive $b_i$ values.
From these results, we can see that changing the $b_i$ distribution essentially only modifies the capacity of the network to experience high welfare; there is little qualitative effect on the
comparative efficacies of the interventions. Across all experiments and heuristics parameterized by $K$, the welfare of the network increases as $K$ grows. This can possibly be explained by the increasing marginal cost of shifting agents’ value parameters. Each additional unit of expenditure for an agent has a smaller impact on the overall network utility, and so we can achieve better results for the same budget by intervening with a larger number of agents. When $K = N$, all 4 heuristics parameterized by $K$ are equivalent to the uniform intervention, and so we observe their utilities converging as expected. The Random-$K$ intervention performs the worst, which is unsurprising, as it does not leverage any factors of the network structure to improve targeting. The Friends-$K$ intervention provides a marginal improvement towards our interventions which require some knowledge of the network structure, but does not drastically outperform Random-$K$. The Top-$K$ and PageRank-$K$ interventions obtain quite similar results. While PageRank and degree count can vary drastically in certain networks, this suggests that they are typically quite correlated in random graphs. As such, the ability to estimate degree counts may be quite useful in approximating PageRank values for a social network. This is not entirely surprising; in a network such as Twitter, the most influential figures are typically those with a large number of followers. Perhaps the most interesting result of these experiments is the relatively miniscule gap between the uniform and optimal interventions. This is quite promising for practical interventions in scenarios which can be approximately captured by the Galeotti et al. model, as it effectively shows that an intervention requiring zero knowledge of network structure can result in nearly as strong of an improvement as the best possible intervention requiring full knowledge of the network structure. As a corollary, this implies that the optimal utility for $C$ can be achieved exactly by applying the uniform intervention with a slightly larger budget $C'$.

Our next experiment will show this more directly.

Varying $C$

We now demonstrate the impact of increasing the budget for intervention. Again, our outcome variable of interest is the total network utility, or $\sum_i w_i$. As the impact of the $b_i$ distribution is merely one of scale, here we limit our results to the case where weights are drawn from a normal distribution. The results shown in Figure 4 are created by fixing $K = 20$ and varying $C$ from 0 to 1000.
The comparative efficacies of each intervention reflect the results from the previous experiments at the points where $K = 20$. By varying the value of $C$, we can understand the growth factor of the efficacy of interventions. This factor appears to be approximately some constant multiple of $\sqrt{C}$, which can be further explained by the increasing marginal cost of intervention. At any level of network utility obtained by the optimal intervention with some budget $C$, plotting a horizontal line at this point gives us a budget $C'$ for each heuristic intervention which matches the resulting utility. This increase in budget will typically be quite small for the uniform intervention, but is much larger for the interventions which can only target a fraction of the network. Thus, our results suggest that the primary bottleneck in effective, budget-constrained interventions is the number of agents we are allowed to target, rather than the amount of information we have about the network structure.

Conclusions and Future Directions

At a high level, our results indicate that effective heuristics exist for targeted interventions in the Galeotti et al. model which require no information about the edges in the network. Additionally, the order of agent degree counts appears to be a close approximation of the ordering of agent PageRank values in random graphs, and random searching as in the Friends-K intervention can be useful in finding high degree vertices. More sophisticated random walk approaches may close this cap even further. It is important to keep in mind the limitations of the model as well as its assumptions about agent behavior. In particular, many social norms may not be easily describable by a real-valued effort level, and would be better described by a binary model; the model and results from Bicchieri and Funcke [3] are more relevant in this case. The increasing marginal cost for both shifting an agent’s preferences and the effort exerted by agents are quite strong assumptions as well, but are perhaps justifiable in
particular cases where the behavior can be appropriately described as the expenditure of
time, labor, or money. Some of the details of the model were selected because they permit
precise theoretical analysis in the full information case. If we are only concerned with expected
performance on random graphs, we might consider relaxations of the model where the power
terms are allowed to vary, possibly mirroring practical scenarios more closely. Despite these
limitations, the qualitative findings may still useful designing heuristic interventions for
scenarios which are not perfectly captured by the model. In some cases, the easiest way
to shift an agent’s utility function for action may be to pay someone in accordance with
the level of action they take. Our results would suggest that we need not necessarily spend
exhaustive efforts mapping the social network in order to successfully incentivize agents to
improve the network utility. Particularly if there is a monetary cost associated with gathering
information, a uniform intervention where all agents are offered a fixed amount of money
to change behavior may be more effective than identifying the most influential agents and
offering them larger sums. Conducting simulations of networks and normative behavior can
be a powerful tool for understanding both quantitative and qualitative dynamics of behavior
in cases when precise theoretical analysis is cumbersome or intractable. While theoretical
guarantees for all networks provide notions of robustness, most real world social networks
are not extreme edge cases, and are well-described by some random distribution of edges.
Even in lieu of a perfect model which encapsulates agent behavior, we might be able to select
effective interventions by finding ones that perform well in simulations for a wide variety of
behavioral models. The efficacy of the uniform intervention suggests that simplicity might
actually take us quite far.

References

[1] Mohammad Akbarpour, Suraj Malladi, and Amin Saberi. Diffusion, seeding, and the


