Altruism Design in Networked Public Goods Games

Sixie Yu[1], **David Kempe**[2], **Yevgeniy Vorobeychik**[1]
[1]Washington University in St. Louis
[2]University of Southern California
[1]{sixie.yu,yvorobeychik}@wustl.edu, [2]david.m.kempe@gmail.com

Abstract

Many collective decision-making settings feature a strategic tension between agents acting out of individual self-interest and promoting a common good. These include wearing face masks during a pandemic, voting, and vaccination. Networked public goods games capture this tension, with networks encoding strategic interdependence among agents. Conventional models of public goods games posit solely individual self-interest as a motivation, even though altruistic motivations have long been known to play a significant role in agents' decisions. We introduce a novel extension of public goods games to account for altruistic motivations by adding a term in the utility function that incorporates the perceived benefits an agent obtains from the welfare of others, mediated by an altruism graph. Most importantly, we view altruism not as immutable, but rather as a lever for promoting the common good. Our central algorithmic question then revolves around the computational complexity of modifying the altruism network to achieve desired public goods game investment profiles. We first show that the problem can be solved using linear programming when a principal can fractionally modify the altruism network. While the problem becomes in general intractable if the principal's actions are all-or-nothing, we exhibit several tractable special cases.

1 Introduction

Individuals in a collective decision-making environment often experience the following type of scenario. Each individual can decide whether or how much effort to invest for the common good; many others may benefit from the efforts, but the cost of the investment is incurred by the individual. Examples of such scenarios include decisions whether or not to wear a mask in a pandemic, vaccinate, or invest in security. The outcomes of such scenarios are often highly suboptimal from a societal point of view: mask-wearing suggestions are flaunted, societies remain undervaccinated, and security measures are not taken. At the heart of this breakdown is that while individuals are "connected" in the sense that their actions affect one another, they are often disconnected "socially," in the sense that they do not experience the utility gain/loss of those affected by their actions. In economic terms, actions have *externalities* on other players, which are, by definition, not *internalized*.

Indeed, the outcomes of such scenarios tend to be significantly different when the individuals form a more tightly knit community. Within families, groups of friends, or small villages, individuals frequently take actions, at a cost to themselves, which primarily benefit others. Similarly, societies with a stronger sense of "duty" towards fellow citizens tend to witness more compliance in all of the above-mentioned examples. Not surprisingly then, campaigns to encourage individual effort (e.g., "Wear a mask — save a life!") tend to appeal to notions of altruism and duty, attempting to get individuals to internalize some of their externalities, if only psychologically.

If the goal of campaigns is to encourage altruistic behavior, an important question is what type of campaign is most effective. Should a principal, aiming to achieve a societally desirable outcome, try to appeal to a generic sense of "duty towards your fellow citizens," try to strengthen the social ties within a small neighborhood, or focus on building a few strong ties between some key individuals? Can the question of how best to *build* or *change* altruism in a society be approached algorithmically, and are the resulting questions tractable or intractable? This is the high-level question we investigate in the present paper.

The question of how to build altruism networks is meaningful in a variety of strategic settings. We focus on *networked public goods games* [Bramoullé and Kranton, 2007; Bramoullé *et al.*, 2014; Feldman *et al.*, 2013; Yu *et al.*, 2020], motivated by the real-world scenarios discussed earlier (e.g., encouraging mask wearing). In networked public goods games, the benefits of an individual's effort are reaped by those with whom the individual interacts, encoded by a network on the individuals.[1] Specifically, an individual's utility depends on 1) her own investment decision, and 2) the *aggregate* investment from her direct neighbors in the network.

In most conventional models of games, including public goods games, it is assumed that agents are driven

[1]Public goods games can be viewed as the special case in which the network is complete.

solely by their individual interests. This assumption is nearly always violated in behavioral studies of public goods games [Ledyard, 1997; Levine, 1998]. While there are many different ways to model altruistic behavior, one natural way was proposed by Ledyard [1997]: the utility of a player i is a linear combination of an *egocentric utility* term, which is the direct benefit to i, and an *altruistic* term, which is a sum of egocentric utilities of other players j, weighted by the strength $a_{i,j}$ of altruism that i feels for j. In most prior work of this kind, $a_{i,j}$ was modeled as a constant for all players i and j. A more general variation by Meier *et al.* [2008] considered an altruism network in vaccination games, but assumed that the altruism graph is identical with the graph representing strategic dependence, as well as that altruism weights are identical for all edges. Naturally, many settings call for more fine-grained models in which the weights can be different: for example, parents typically care more about their children's welfare than that of strangers.

While the focus of past work on altruism in games has been on its equilibrium effects, our point of departure is to consider the altruism network itself as (partially) under the control of a principal. In other words, we view altruistic motivations as a lever that can be adjusted to promote the common good, for example, through public outreach campaigns, community meetings, and personal introductions. Specifically, we propose a model of modifying altruism networks, with the goal of inducing a target investment profile by the agents. We consider three variants of the altruism network: weighted, directed, and undirected. We show that even for very complex available actions, the problem can be solved efficiently using linear programming when the principal has fine-grained control over the extent to which actions are taken. When the principal can only control *which* actions are taken, the problem becomes NP-complete, even when each action affects only a single edge in the altruism network. However, when the altruism network is directed, we show that the problem is tractable in a broad array of special cases by reductions to the (tractable cases of) the KNAPSACK problem. We also leverage this connection to exhibit an FPTAS for the general case. When the altruism network is undirected, the hardness results apply even for much more restrictive special cases. However, we show that the problem is tractable when the benefits from investment are linear and uniform, by a non-trivial reduction to the problem of NETWORK DESIGN FOR DEGREE SETS (NDDS) introduced by [Kempe *et al.*, 2020], who also showed that it can be solved in polynomial time.

Our problem of designing an altruism graph to achieve target equilibrium outcomes is, indeed, conceptually related to Kempe *et al.* [2020], who study the problem of designing the *strategic* network in networked public goods games. The main rationale for shifting focus to designing *altruism graphs* is that strategic networks are often difficult to change. For example, in a pandemic, it is difficult to directly affect contacts among individuals, as these are ultimately the products of individual choices (e.g., even lockdowns may be ineffective if individuals are non-compliant, except through levels of enforcement that are often viewed as unacceptable by the population). In contrast, it can be significantly easier to try to impact decisions indirectly by evoking altruistic motivations in people. From a technical perspective, the problem of altruism design impacts utilities linearly, in contrast to the design of strategic networks; however, it is also distinct from the linear special case in Kempe *et al.* [2020], where the marginal impact of each neighbor on a player's utility is identical, in contrast to altruism design, where these differ.

Related Work Our work is related to four lines of research: graphical games, altruism modeling, mechanism and market design, and network design. Graphical games encode sparsity in the interdependence of player utility functions using a graph [Kearns *et al.*, 2001; Shoham and Leyton-Brown, 2008], with networked public goods games an important class of such models [Bramoullé and Kranton, 2007; Galeotti *et al.*, 2010; Grossklags *et al.*, 2008; Yu *et al.*, 2020]. A conventional assumption in such games is that agents act to exclusively promote their own interest. However, considerable experimental evidence exists that even games with this payoff structure elicit altruistic motivations among human subjects [Dong *et al.*, 2016; Levine, 1998]. This, in turn, led to a series of approaches to model altruism in a variety of games, including public goods games, which are of direct interest here [Ledyard, 1997; Dong *et al.*, 2016], inoculation games [Meier *et al.*, 2008], routing games [Chen and Kempe, 2008], and congestion games [Chen *et al.*, 2011, 2014].

A typical way that altruism is captured in prior literature is by either adding a social welfare term to utility functions [Ledyard, 1997], or introducing a parameter that governs the extent to which agents care about their social network neighbors [Meier *et al.*, 2008]. Our model is distinct in that it allows altruism to be relationship-dependent, a property we model by an altruism network. Moreover, our goal is to *modify* an altruism network to achieve a target equilibrium (e.g., one that maximizes social welfare).

Mechanism and market design also aim to change the parameters of a game to induce desirable equilibrium outcomes [Nisan *et al.*, 2007; Haeringer, 2018; Dughmi, 2017]. We introduce altruism network design as a novel lever for aligning incentives with public good.

The last relevant line of research is network design. A particularly related thread in network design is to study the effects of network modification on equilibrium outcomes or welfare [Kempe *et al.*, 2020; Bramoullé and Kranton, 2007; Galeotti *et al.*, 2010]. Another related thread is to alter a network in order to effect a variety of different outcomes for different types of games [Sheldon *et al.*, 2010; Chen *et al.*, 2016; Ghosh and Boyd, 2006; Tong *et al.*, 2012; Bredereck and Elkind, 2017; Sina *et al.*, 2015; Matteo Castiglioni, 2020; Amelkin and Singh, 2019; Garimella *et al.*, 2018].

2 Networked Public Goods Games and Altruism

We study altruism in *binary networked public goods games (BNPGs)*, which are an important variant of *public goods games* studied extensively in prior literature [Bramoullé and Kranton, 2007; Galeotti *et al.*, 2010; Grossklags *et al.*, 2008; Suri and Watts, 2011; Dong *et al.*, 2016; Kempe *et al.*, 2020; Yu *et al.*, 2020]. We begin by formally describing BNPGs and pure strategy Nash equilibria, the solution concept we focus on. We then discuss a natural model of altruism in games, and its application to the specific case of BNPGs.

Binary Networked Public Goods Games: A *binary networked public goods* (BNPG) game is characterized by the following:

1. A simple, undirected, and loop-free graph $H = (V, E_H)$ in which the nodes $V = \{1, 2, \dots, n\}$ are the agents/players, and the edges E_H represent the interdependencies among the players' payoffs.
2. A binary strategy space $\{0, 1\}$ for each player i. [2] We interpret the choice of strategy 1 as *investing* in a public good, while choosing 0 is interpreted as non-investment. The action of player i is denoted by x_i, and the joint pure strategy profile of all players by $\boldsymbol{x} = (x_1, x_2, \dots, x_m)$. We use $\boldsymbol{x}_{-i}$ to denote a strategy profile that omits player i's strategy.
3. A non-decreasing utility function $U_i(x_i, \boldsymbol{x}_{\mathcal{N}_i^{(H)}})$ for each player i, where $\mathcal{N}_i^{(H)} = \{j \mid (i, j) \in E_H\}$ is the set of i's neighbors in the graph H.

As is common in the literature on public goods games [Bramoullé and Kranton, 2007], we assume that each player's (egocentric) utility function U_i only depends on the total investment by i's network neighbors. To formalize this, we define $n_i^{(H,\boldsymbol{x})} = \sum_{j \in \mathcal{N}_i^{(H)}} x_j$ as the number of i's neighbors who invest under $\boldsymbol{x}$. We omit H, $\boldsymbol{x}$, or both from this notation when they are clear from the context. Each player i's utility function then has the following form:

$$U_i(\boldsymbol{x}) = U_i(x_i, n_i^{(\boldsymbol{x})}) \;=\; g_i(x_i, n_i^{(\boldsymbol{x})}) - c_i x_i. \tag{1}$$

The second term ($-c_i x_i$) captures the cost incurred by player i from investing. As is standard in the public goods games literature, each g_i is assumed to be a non-negative and non-decreasing function of both of its arguments, capturing the positive externality that i experiences from her neighbors' (and her own) investments. Observe that each function g_i can be represented using $O(n)$ values, so the entire BNPG game (including the graph structure) can be represented using $O(n^2)$ values.

We will consider *pure-strategy Nash equilibria (PSNE)* of BNPGs. A pure strategy profile $\boldsymbol{x}^*$ is a PSNE if for all i, $U_i(x_i, n_i^{(H,\boldsymbol{x})}) \geq U_i(1 - x_i, n_i^{(H,\boldsymbol{x})})$. We write $\mathcal{E}(\mathcal{G})$ for the set of all PSNEs of the game $\mathcal{G}$.

Altruistic Motivations in BNPGs: A natural way to model other-regarding utilities is to define a player's utility as a linear combination of her egocentric utility, defined by Equation (1), and the egocentric utilities of other players.

To formalize this, we can think of the matrix $A = (a_{i,j})_{i,j}$ as encoding an *altruism network*. This captures the central motivation of our work, discussed in the introduction: that the agents who are *affected* by the actions of i may not be the same as the agents that i *cares about*.[3] The resulting utility function of a player i in our BNPG game model with altruism is

$$U_i^{(A)}(\boldsymbol{x}) = g_i(x_i, n_i) - c_i x_i + \sum_{j \in \mathcal{N}_i^{(H)}} a_{i,j} g_j(x_j, n_j). \tag{2}$$

We denote the BNPG with altruism network A by BNPG(A).

We note two points about this model: First, the altruistic term of player i's utility does not include a term for the *investment cost* of player j, only the *utility*. This is inconsequential, as investment decisions by j are not under i's control, so from i's point of view, these cost terms are constants. Second, the other-regarding terms have no component in which j's utility due to i's egocentric payoff is recursively considered. Such utilities may be harder to observe, and from a modeling perspective, they can be transformed into the case we study here; see [Bergstrom, 1999].

In some of the results in later sections, we will specifically want to stress the *network* aspect of altruism. In that case, we will assume that we are given a (directed or undirected) *altruism graph* G, and that $a_{i,i} = 1$ for all i, $a_{i,j} = a$ for all i, j for which G contains the edge (i, j), and $a_{i,j} = 0$ for all i, j for which G contains no edge. In other words, the altruism strength of the edges of G is uniform. When G is undirected, we will refer to the case as *symmetric altruism*; when G is directed, we call it *asymmetric altruism*.

3 Modifying Altruism Networks

As discussed in the introduction, a major problem with public goods situations is that equilibria can be far from optimal because individuals may not fully internalize the impact of their actions on others. Therefore, a principal who seeks to steer the network to a better equilibrium might wish for agents to consider others in their decisions. We consider situations in which the principal can increase or decrease the salience of others in these settings, for example, through introductions, advertising, or community meetings.[4] We now formally model this problem as modifying an altruism network to achieve socially desired outcomes.

Our model is as follows. The principal aims to induce a particular *target investment profile* $\boldsymbol{x}^*$. This allows us to cleanly capture a broad variety of design goals, such as maximizing welfare or achieving fairness, while focusing on the computational issues at the core of our specific problem of altruism design. To induce the outcome, the principal wants to (minimally) modify the altruism network A^{in} to A such that $\boldsymbol{x}^*$ is a PSNE of the modified game BNPG(A).[5]

[2] The public goods game literature extensively considers both binary and continuous decisions, and both are natural candidates for considering network design. Here, we focus on binary strategies due to their applicability to decisions such as vaccinations or mask wearing.

[3] In the context of vaccination games, Meier *et al.* [2008] study the special case in which the friendship network is the *same* as the network of who may transmit a disease to whom.

[4] This is separate from, and in addition to, other channels, such as rewarding or punishing certain actions.

[5] There may be other PSNE of the game. We implicitly assume that the principal can *suggest* an equilibrium to the agents, who will follow the suggestion unless it is in their best interest to deviate.

As implied by the preceding discussions, the principal may have at his disposal a number of different actions, affecting the altruism between different sets of pairs of agents, in positive or negative ways. For example, a general appeal to watch out for one another may lead to a small increase in altruism between many pairs of individuals; a community meeting may lead to a stronger increase among a smaller subset, and a personal introduction may introduce one strong edge. We model such settings by assuming that there are K *actions*, with K polynomial in n. Each action k has associated with it a set $\mathcal{S}^{(k)}$ of affected altruism edges, a cost $\gamma^{(k)} \geq 0$, and a *sign* $\sigma^{(k)} \in \{-1, 1\}$, which captures whether the action strengthens or weakens the edges in $\mathcal{S}^{(k)}$.[6] The edge set is encoded in the corresponding adjacency matrix $M^{(k)}$, with entries $m_{i,j}^{(k)}$ which are 1 for $(i,j) \in \mathcal{S}^{(k)}$ and 0 otherwise. The costs $\gamma^{(k)}$ measure the monetary expense or effort/time needed to implement the corresponding activity, per unit of change in the altruism. The principal aims to solve the following problem:

Definition 3.1 (Altruism Network Modifications (ANM))**.** ***Given:*** *an altruism network A^{in}, target investment profile $\boldsymbol{x}^*$, and actions $\{\mathcal{S}^{(1)}, \ldots, \mathcal{S}^{(K)}\}$ with signs $\sigma^{(k)}$ and costs $\gamma^{(k)}$.* ***Goal:*** *choose a non-negative vector $\boldsymbol{v} \in \mathbb{R}^K_{\geq 0}$ of minimum total cost $\sum_{k=1}^K v_k \gamma^{(k)}$ such that the modified game with new altruism network*

$$A = A^{in} + \sum_{k=1}^{K} v_k \cdot \sigma^{(k)} \cdot M^{(k)} \tag{3}$$

has $\boldsymbol{x}^*$ as a PSNE, i.e., $\boldsymbol{x}^* \in \mathcal{E}(\mathrm{BNPG}(A))$.

Here, $\boldsymbol{v}$ is a vector capturing how much the principal spends on each of the available actions. We assume that the different actions are cumulative in their effects on any of the network's edges, and (partially) cancel out when they have opposite signs. Note that we could additionally truncate the entries of A so that $0 \leq A \leq 1$; this is not consequential for our results, and we proceed with the slightly cleaner model above. The principal's spending on actions results in a modified altruism network, and his goal is to ensure that the target action profile $\boldsymbol{x}^*$ becomes one of the equilibria of the modified game. The principal wants to achieve this goal at minimum cost (which is infinite if the problem is infeasible).

Since strategies are binary, a target profile $\boldsymbol{x}^*$ can be equivalently represented by the set of agents who invest under this profile, $\mathcal{I} = \mathcal{I}_{\boldsymbol{x}^*} = \{i \in V \mid x_i^* = 1\}$. Whether or not players invest can be completely characterized using a collection of inequalities. In these inequalities, what ultimately determines the decision is the agent's marginal value from investing. This marginal value has two elements: first, the agent's own marginal benefit, $\Delta_{x_i} g_i(n_i) := g_i(1, n_i) - g_i(0, n_i)$, and second, the marginal benefit that i can obtain from altruism towards j, which is $\Delta_j^- := g_j(x_j, n_j) - g_j(x_j, n_j - 1)$ for agents i who invest, and $\Delta_j^+ := g_j(x_j, n_j + 1) - g_j(x_j, n_j)$ for those agents i who do not. We further define $\theta_i := c_i - \Delta_{x_i} g_i(n_i)$ for each agent i. We then obtain the following equivalent characterization of a PSNE of $\mathrm{BNPG}(A)$, in which a set $\mathcal{I}$ of players invest, and the rest do not:

$$\begin{aligned} \sum_{j \in \mathcal{N}_i^{(H)}} a_{i,j} \Delta_j^- &\geq \theta_i \quad \text{if } i \in \mathcal{I} \\ \sum_{j \in \mathcal{N}_i^{(H)}} a_{i,j} \Delta_j^+ &\leq \theta_i \quad \text{if } i \notin \mathcal{I}, \end{aligned} \tag{4}$$

with $a_{i,j}$ the strength of i's altruism for j under A.

The marginal benefit functions $\Delta g_i := (\Delta_i^-, \Delta_i^+)$ are important parameters in our model; they can be restricted by putting limitations on g_i, e.g., Δg_i is bounded by a polynomial in n iff g_i itself is. We consider three possible restrictions:

general g_i: Δ_i^- and Δ_i^+ are arbitrary.

polynomial g_i: When the g_i are bounded by a polynomial in n, Δ_i^- and Δ_i^+ are also bounded by a polynomial in n.

uniform linear and separable g_i: When all g_i are of the form $g_i(x_i, n_i) = h_i(x_i) + \Delta \cdot n_i$ for some possibly idiosyncratic function h_i and some (common) constant Δ, then $\Delta_i^- = \Delta_i^+ = \Delta$ for all i and all values of n_i.

4 An LP for Fractional Modifications

We begin by considering the variant of the problem in which the principal can spend fractionally on each action, i.e., $\boldsymbol{v} \in \mathbb{R}^K_{\geq 0}$. In that case, the principal's optimal strategy can be found using a straightforward linear program. The decision variables are simply the principal's investments $v_k \geq 0$. For ease of notation, we also define variables $a_{i,j}$ for the altruism from i towards j resulting from the modifications. Given the available actions and costs, the relationship between $a_{i,j}$ and v_k is exactly characterized by Equation (3), which is linear in the v_k. Next, notice that given the BNPG and the desired PSNE $\boldsymbol{x}^*$, the set $\mathcal{I}$ and all relevant constants in Equation (4) (that is $\Delta_i^-, \Delta_i^+, \theta_i$) can be immediately computed. Therefore, we obtain the following linear program for finding the optimal spending strategy for the principal:

$$\begin{array}{ll} \text{Minimize } \sum_{k=1}^K v_k \cdot \gamma^{(k)} \text{ subject to} & \\ a_{i,j} = a_{i,j}^{\text{in}} + \sum_{k=1}^K v_k \cdot \sigma^{(k)} \cdot m_{i,j}^{(k)} & \text{for all } i, j \\ \sum_{j \in \mathcal{N}_i^{(H)}} a_{i,j} \Delta_j^- \geq \theta_i & \text{for all } i \in \mathcal{I} \\ \sum_{j \in \mathcal{N}_i^{(H)}} a_{i,j} \Delta_j^+ \leq \theta_i & \text{for all } i \notin \mathcal{I} \\ v_k \geq 0 & \text{for all } k. \end{array} \tag{5}$$

Notice that we could fairly straightforwardly generalize the LP to even deal with more general actions, namely, allowing an action k to affect different edges in different ways by allowing $M^{(k)}$ to have entries that are not all equal.

5 Graph-Based Modifications

While there are contexts in which a principal can precisely control the amount of effort invested in different actions, there are many others in which actions take more of an all-or-nothing nature. Indeed, many activities that increase altruism through making harms to others more salient, such as community meetings or public health advertising campaigns (e.g.,

[6]Typically, a principal would be more likely to want to *strengthen* the altruism network. However, one can easily imagine situations and construct instances in which a weakening of the network is necessary. We therefore aim for more generality in our model.

to promote mask-wearing), are naturally discrete (a 1-second ad is not very effective) and, thus, have a discrete impact on the altruism graph. This motivates the special case in which all entries of $\boldsymbol{v}$ must be binary, corresponding to decisions whether or not the principal will add/remove the edge sets $\mathcal{S}^{(k)}$. As we show presently that even if we restrict $\mathcal{S}^{(k)}$ to affect a single edge the problem now becomes NP-hard, we devote the sequel to this special case, and seek to identify what additional structure is sufficient to make the problem, or its approximation, tractable.

When $\mathcal{S}^{(k)}$ is a singleton (i, j), we can think of the input as a graph $G^{\text{in}} = (V, E^{\text{in}})$ (instead of a weighted network). There is a cost $\gamma^{(i,j)}$ associated with each (directed or undirected) node pair (i, j). If $(i, j) \in E^{\text{in}}$, then this is the cost for removing (i, j); otherwise, it is the cost of adding (i, j). Thus, implicitly, $\sigma^{(i,j)} = 1$ if $(i, j) \notin E^{\text{in}}$, and $\sigma^{(i,j)} = -1$ if $(i, j) \in E^{\text{in}}$. Adding/removing edges results in a new altruism network $G = (V, E)$. All off-diagonal non-zero entries of the altruism network A then have the same altruism value a, while all diagonal entries are set to 1.

We study two variants of this problem: 1) asymmetric altruism, that is, when the altruism graph is directed, and 2) symmetric altruism, where it is undirected. Typically, both will capture some important aspect of the real world: while altruism often aligns with actual social or kinship ties, it can also result from a general sense of responsibility or goodwill, which may not be reciprocated.

Asymmetric Altruism

We begin by formally showing that in this setting, ANM is in general intractable, even in the special case where we can only add or remove individual edges, rather than subsets of edges.

Theorem 5.1. *ANM with asymmetric altruism is NP-complete even when:*

1. *the sets $\mathcal{S}^{(k)}$ are singletons and can only be added, i.e., $\sigma^{(k)} = 1$ for all k,*
2. *the initial altruism network is empty, and*
3. *the target profile $\boldsymbol{x}^*$ has all agents investing.*

The proof is a direct reduction from the KNAPSACK problem and given in the Supplementary Material.

Next, we show that under mild additional assumptions, the problem becomes efficiently solvable. The key observation that enables our positive results is that for directed graphs, adding or removing an edge $i \to j$ only affects agent i's altruistic behavior. This allows us to decompose the problem into n independent subproblems only connected through a common budget constraint. Each subproblem can be naturally modeled as a KNAPSACK problem, and so long as the KNAPSACK problems are individually solvable, so is the overall problem. More specifically, we distinguish two cases, based on whether $i \in \mathcal{I}$.

- If $i \in \mathcal{I}$, then the altruism edges originating with i must ensure that $\sum_{j \in \mathcal{N}_i^{(H)}} a_{i,j} \Delta_j^- \geq \theta_i$. Let $\phi_i = \sum_{j \in \mathcal{N}_i^{(H)}} a_{i,j}^{\text{in}} \Delta_j^-$. If $\phi_i \geq \theta_i$, then i will invest even without adding any edges. Otherwise, the principal will need to add edges out of i adding a total altruism term of $\theta_i - \phi_i$; also, the principal will never want to remove edges out of i. Adding a directed edge $i \to j$ can be thought of as putting an item with value $a \cdot \Delta_j^-$ and cost/weight $\gamma^{(i,j)}$ into a knapsack. Thus, the set of "items" available to add to agent i is $\mathcal{P}_i = \{j \in \mathcal{N}_i^{(H)} \mid a_{i,j}^{\text{in}} = 0\}$. The subproblem is then to select items from $\mathcal{P}_i$ such that the total value is at least $\theta_i - \phi_i$ while the total weight is minimized.
- Similarly, if $i \notin \mathcal{I}$, then the altruism edges originating with i must ensure that $\sum_{j \in \mathcal{N}_i^{(H)}} a_{i,j} \Delta_j^+ \leq \theta_i$. Let $\phi_i = \sum_{j \in \mathcal{N}_i^{(H)}} a_{i,j}^{\text{in}} \Delta_j^+$. Analogously to the previous case, the principal now wants to *remove* edges such that the altruism is reduced by at least $\phi_i - \theta_i$ (unless $\phi_i \leq \theta_i$, in which case nothing needs to be done). Again, this problem can be modeled as a KNAPSACK problem. The set of items available is $\mathcal{P}_i = \{j \in \mathcal{N}_i^{(H)} \mid a_{i,j}^{\text{in}} = 1\}$. The directed edge $i \to j$ is modeled as an item with value $a \cdot \Delta_j^+$ and cost/weight $\gamma^{(i,j)}$. The goal is to minimize the total weight subject to achieving total value at least $\phi_i - \theta_i$.

It is well known [Kleinberg and Tardos, 2005; Vazirani, 2001] that the KNAPSACK problem can be solved in polynomial time using Dynamic Programming when either the weights or the values are bounded by a polynomial in the number of items. Using standard rounding/scaling techniques, this approach also yields an FPTAS. By leveraging these algorithms, we obtain the corresponding results for our problem.

General g_i and Polynomial Edge Costs When edge costs are bounded by a polynomial in n, which corresponds to polynomially bounded item weights in the KNAPSACK instances, which are therefore polynomial-time solvable with Dynamic Programming (DP). Applying DP for each agent i separately then yields a minimum-cost overall solution. We obtain the following:

Proposition 5.2. *Under asymmetric altruism, the problem ANM with general g_i and polynomially bounded edge costs is polynomial-time solvable.*

Polynomial g_i and General Edge Costs When all the g_i are polynomially bounded, so are their differences, and hence the (scaled) item values Δ_j^+ and Δ_j^-. Hence all the values of the "items" are polynomially bounded. Tractability then follows from the fact that the KNAPSACK problem is polynomial-time solvable (using Dynamic Programming) when item values are bounded by some polynomial in the length of the input. Again, this allows an algorithm to solve each subproblem in polynomial time, and then aggregate the optimal solutions. This gives rise to the following proposition:

Proposition 5.3. *Under asymmetric altruism, the problem ANM with polynomially bounded g_i and general edge costs is polynomial-time solvable.*

An FPTAS for the general case Finally, we can leverage the standard FPTAS for KNAPSACK to obtain an FPTAS for ANM with general asymmetric altruism. Specifically, given a parameter ϵ, one can run the FPTAS with that parameter for each of the subproblems/agents i separately. The result for each i will be a set of edges to add/remove such that i invests iff $x_i^* = 1$, and the total cost of the modifications is within a factor $(1+\epsilon)$ of optimal. Adding all of these costs shows that the overall cost is within a factor $(1+\epsilon)$ of optimal. We obtain the following proposition:

Proposition 5.4. *Under asymmetric altruism, consider ANM with general g_i and general edge costs. Given $\epsilon > 0$, the optimal cost B^* can be approximated arbitrarily well, i.e., a solution of cost $B \leq (1+\epsilon)B^*$ can be found, with an algorithm which runs in time polynomial in n and $1/\epsilon$.*

Symmetric Altruism

Next, we turn to the setting in which altruism is reciprocal or symmetric: when an edge is added to the altruism network, it affects *both* incident agents. While many graph-theoretic questions are easier for undirected graphs than for directed ones, the ANM problem becomes harder. Intuitively, the reason is that while adding the edge (i, j) is very beneficial for i, it may be less so for j; given the choice, adding a different edge out of j may be preferable. Under the symmetric altruism model, the principal does not have the fine-grained control of adding different edges out of i and j, and might have to "waste" one direction of the edge. The resulting "side-effects" of desirable edges must be more globally balanced. Indeed, we show that even special cases that are polynomial-time solvable in the asymmetric model become NP-hard in the symmetric model.

Theorem 5.5. *ANM with symmetric altruism is NP-complete even when*
- *all the sets $\mathcal{S}^{(k)}$ are singletons,*
- *all agents invest under the target equilibrium $\boldsymbol{x}^*$,*
- *all g_i are polynomially bounded,*
- *all edge costs $\gamma^{(k)}$ are 1, and*
- *the graph H is a clique.*

The proofs of this and the remaining results are in the Supplementary Material. Recall that for the asymmetric case, even just one of uniform (or even just polynomially bounded) edge costs and polynomially bounded g_i is enough to obtain tractability.

For the remainder of this section, we will focus on the special case when all agents invest under the target equilibrium $\boldsymbol{x}^*$. This is because for more general target equilibria, even deciding if there exists *any* altruism graph yielding this equilibrium is NP-complete. In other words, even the decision problem of a principal with infinite budget is NP-complete.

Theorem 5.6. *ANM with symmetric altruism for arbitrary target equilibria $\boldsymbol{x}^*$ is NP-complete even when*
- *all the sets $\mathcal{S}^{(k)}$ are singletons,*
- *all g_i are polynomially bounded,*
- *all edge costs $\gamma^{(k)}$ are 0 (i.e., the principal's budget is infinite),*
- *the graph H is a clique.*

Theorem 5.6 implies that no approximation guarantee can be attained in polynomial time for ANM when $\boldsymbol{x}^*$ is an arbitrary action profile. However, we remark that when all agents invest under $\boldsymbol{x}^*$, there is a straightforward polynomial-time $(2+\epsilon)$-approximation algorithm: the algorithm applies the FPTAS to the asymmetric version and adds a reciprocal edge whenever a directed edge is added. This leads to a blowup of a factor 2 compared to the FPTAS achieved in the asymmetric setting.

Uniform Separable Linear Utility Functions When the agents have utility functions g_i that are separable in the arguments, and linear with common slope in the second argument (we term these uniform separable and linear, or USL), the marginal benefits are uniform and equal to a constant Δ, i.e., $\Delta_i^- = \Delta_i^+ = \Delta$ for all i and all values of n_i. Such utility functions are commonly studied in public goods games [Suri and Watts, 2011]. We show that when the goal is for all players to invest, for USL utility functions and when all sets $\mathcal{S}^{(k)}$ are single edges, the problem becomes tractable, even with all other parameters (edge costs, network structure, etc.) being fully general.

Theorem 5.7. *ANM with symmetric altruism is polynomial time solvable when*
- *the utility functions g_i are USL,*
- *the sets $\mathcal{S}^{(k)}$ are singleton,*
- *the target equilibrium $\boldsymbol{x}^*$ has all agents investing.*

The proof of Theorem 5.7 (given in the Supplementary Material) is through a non-trivial reduction to the MIN-COST PERFECT MATCHING problem via a connection to another related problem: NETWORK DESIGN FOR DEGREE SETS (NDDS) [Kempe *et al.*, 2020].

6 Conclusion

We consider how to change altruistic behavior of individuals so as to induce societally desirable outcomes. One major contribution of our work is to separately capture the strategic interdependencies and the altruistic behaviors of individuals. We propose a model of modifying the altruism network, with the goal of inducing a target investment profile by the individuals. A series of corresponding algorithmic results are exhibited, including hardness results even in very restrictive scenarios (e.g., each modification only affects a single edge), and tractability results in a broad array of special cases.

Our work only focused on the goal of achieving a *specific* strategy profile $\boldsymbol{x}^*$. Other natural goals would be to maximize the social welfare of the resulting profile (e.g., subject to a budget constraint), or to maximize the number of investing players. In these cases, many of our current reductions will fail, since they rely on knowing exactly which set of players will invest. More fundamentally, the simultaneous study of externality and altruism networks raises interesting structural questions. Is there a sense in which for some broad class of games (such as public goods games), it is desirable for these networks to be aligned? In other words, should individuals care most about those who are most affected by their actions? Or can it be beneficial to have "long-range" edges? Additionally, our model of altruism assumes that edges can be added/removed by the principal, so long as the principal is willing to pay for the action. A natural alternative would be to recognize that the total "capacity" of an agent for altruism towards others may be bounded. In other words, the principal may be able to introduce new altruism edges, but the more outgoing edges agent i has, the weaker each of them becomes. This change will require different algorithmic and structural insights.

References

Victor Amelkin and Ambuj K. Singh. Fighting opinion control in social networks via link recommendation. In *Proc. 25th Intl. Conf. on Knowledge Discovery and Data Mining*, pages 677–685, 2019.

Theodore C. Bergstrom. Systems of benevolent utility functions. *J. of Public Economic Theory*, 1(1):71–100, 1999.

Yann Bramoullé and Rachel Kranton. Public goods in networks. *Journal of Economic Theory*, 135(1):478–494, 2007.

Yann Bramoullé, Rachel Kranton, and Martin D'Amours. Strategic interaction and networks. *American Economic Review*, 104(3):898–930, 2014.

Robert Bredereck and Edith Elkind. Manipulating opinion diffusion in social networks. In *Proc. 26th Intl. Joint Conf. on Artificial Intelligence*, pages 894–900, 2017.

Po-An Chen and David Kempe. Altruism and selfishness in traffic routing. In *Proc. 9th ACM Conf. on Electronic Commerce*, pages 140–149, 2008.

Po-An Chen, Bart de Keijzer, David Kempe, and Guido Schäfer. The robust price of anarchy of altruistic games. In *Proc. 7th Workshop on Internet and Network Economics (WINE)*, pages 383–390, 2011.

Po-An Chen, Bart de Keijzer, David Kempe, and Guido Schäfer. Altruism and its impact on the price of anarchy. *ACM Trans. Economics and Computation*, 2(4):17, 2014.

Chen Chen, Hanghang Tong, B. Aditya Prakash, Tina Eliassi-Rad, Michalis Faloutsos, and Christos Faloutsos. Eigen-optimization on large graphs by edge manipulation. *ACM Transactions on Knowledge Discovery from Data*, 10(4):49:1–49:30, 2016.

Yali Dong, Boyu Zhang, and Yi Tao. The dynamics of human behavior in the public goods game with institutional incentives. *Scientific Reports*, 6(1):1–7, 2016.

Shaddin Dughmi. Algorithmic information structure design: A survey. *ACM SIGecom Exchanges*, 15(2):2–24, 2017.

Michal Feldman, David Kempe, Brendan Lucier, and Renato Paes Leme. Pricing public goods for private sale. In *Proc. 14th ACM Conf. on Electronic Commerce*, pages 417–434, 2013.

Andrea Galeotti, Sanjeev Goyal, Matthew O. Jackson, Fernando Vega-Redondo, and Leeat Yariv. Network games. *The Review of Economic Studies*, 77(1):218–244, 2010.

Michael Garey and David S. Johnson. *Computers and Intractability: A guide to the Theory of NP-completeness*. Freeman, 1979.

Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Reducing controversy by connecting opposing views. In *Proc. 27th Intl. Joint Conf. on Artificial Intelligence*, pages 5249–5253, 2018.

Arpita Ghosh and Stephen Boyd. Growing well-connected graphs. In *Proc. 45th IEEE Conference on Decision and Control (CDC)*, pages 6605–6611, 2006.

Jens Grossklags, Nicolas Christin, and John Chuang. Security and insurance management in networks with heterogeneous agents. In *Proc. 9th ACM Conf. on Electronic Commerce*, pages 160–169, 2008.

Guillaume Haeringer. *Market Design: Auctions and Matching*. The MIT Press, 2018.

Michael Kearns, Michael L. Littman, and Satinder Singh. Graphical models for game theory. In *Proc. 17th Conf. on Uncertainty in Artificial Intelligence*, pages 253–260, 2001.

David Kempe, Sixie Yu, and Yevgeniy Vorobeychik. Inducing equilibria in networked public goods games through network structure modification. In *Proc. 19th Intl. Conf. on Autonomous Agents and Multiagent Systems*, pages 611–619, 2020.

Jon Kleinberg and Eva Tardos. *Algorithm Design*. Addison-Wesley, 2005.

John Ledyard. Public goods: A survey of experimental resesarch. In J. Kagel and A. Roth, editors, *Handbook of Experimental Economics*, pages 111–194. Princeton University Press, 1997.

David Levine. Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics*, 1:593–622, 1998.

Nicola Gatti Matteo Castiglioni, Diodato Ferraioli. Election control in social networks via edge addition or removal. In *Proc. 34th AAAI Conf. on Artificial Intelligence*, 2020.

Dominic Meier, Yvonne Anne Oswald, Stefan Schmid, and Roger Wattenhofer. On the windfall of friendship: Inoculation strategies on social networks. In *Proc. 9th ACM Conf. on Electronic Commerce*, pages 294–301, 2008.

Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay Vazirani, editors. *Algorithmic Game Theory*. Cambridge University Press, 2007.

Daniel Sheldon, Bistra Dilkina, Adam N. Elmachtoub, Ryan Finseth, Ashish Sabharwal, Jon Conrad, Carla Gomes, David Shmoys, William Allen, Ole Amundsen, and William Vaughan. Maximizing the spread of cascades using network design. In *Proc. 26th Conf. on Uncertainty in Artificial Intelligence*, pages 517–526, 2010.

Yoav Shoham and Kevin Leyton-Brown. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, 2008.

Sigal Sina, Noam Hazon, Avinatan Hassidim, and Sarit Kraus. Adapting the social network to affect elections. In *Proc. 14th Intl. Conf. on Autonomous Agents and Multiagent Systems*, pages 705–713, 2015.

Siddharth Suri and Duncan J Watts. Cooperation and contagion in web-based, networked public goods experiments. *PloS One*, 6(3):e16836, 2011.

Hanghang Tong, B. Aditya Prakash, Tina Eliassi-Rad, Michalis Faloutsos, and Christos Faloutsos. Gelling, and melting, large graphs by edge manipulation. In *Proc. 21st ACM Intl. Conf. on Information and Knowledge Management*, pages 245–254, 2012.

Vijay V. Vazirani. *Approximation Algorithms*. Springer, 2001.

Sixie Yu, Kai Zhou, P. Jeffrey Brantingham, and Yevgeniy Vorobeychik. Computing equilibria in binary networked public goods games. In *Proc. 34th AAAI Conf. on Artificial Intelligence*, 2020.

Supplementary Material

A Proof of Theorem 5.1

We consider the decision version of ANM with an input parameter B, the available budget. The goal is to decide if the total cost of producing a graph G with $\boldsymbol{x}^* \in \mathcal{E}(\text{BNPG}(G))$ exceeds B. It is clear that the problem is in NP: given a candidate set of actions, their costs can be added, and one can verify that the desired equilibrium $\boldsymbol{x}^*$ is indeed a PSNE.

To prove NP-harness, we reduce from the KNAPSACK problem. Recall that in the KNAPSACK problem, we are given a knapsack with capacity W, a set of n items, and a value V. Each item has a given weight $w_i \geq 0$ and a value $v_i \geq 0$. The goal is to decide if there is a subset S of items, such that the total weight $\sum_{i \in S} w_i \leq W$ and the total value $\sum_{i \in S} v_i \geq C$.

Given an instance of the KNAPSACK problem, we construct an instance of ANM as follows. The target investment profile $\boldsymbol{x}^*$ is that all players invest. The altruism constant $a > 0$ is arbitrary (e.g., $a = 1$). The graph $H = (V, E_H)$ is a clique with $n+1$ nodes; a special node u_c together with n nodes $u_1, \ldots, u_n$. The initial altruism network G^{in} is empty, i.e., $a^{\text{in}}_{u_i,u_j} = 0$ for all $i \neq j$. For the node u_c, let $\theta_{u_c} = a \cdot C$ and $\Delta^-_{u_c}$ be an arbitrary nonnegative number. For each node u_i, let $\theta_{u_i} = 0$ and $\Delta^-_{u_i} = v_i$. Intuitively, the node u_i corresponds to the item with value v_i.

To realize the above, the function g_{u_c} is constant over the whole domain, i.e., $\Delta_{x_{u_c}} g_{u_c} = 0$ and $\Delta^-_{u_c} = 0$. The cost $c_{u_c} = a \cdot C$. This ensures that $\theta_{u_c} = a \cdot C$. For a node u_i, we let g_{u_i} be a linear function over the first argument with slope c_{u_i}, such that $\Delta_{x_i} g_{x_i} = c_{u_i}$. The function g_{u_i} over the second argument is also a linear function with slope v_i, i.e., $\Delta^-_{u_i} = v_i$.

The cost to add an edge from u_c to u_i is w_i. Neither edges from u_i to u_c nor edges between different u_i can be added, i.e., their costs are infinite. The budget is the weight capacity, $B = W$. The reduction clearly can be constructed in polynomial time.

First, suppose that there is a set S of items that solves the KNAPSACK problem. Then, picking the actions/edges (u_c, u_i) for $i \in S$ does not exceed the budget B. Each node u_i will invest as $\theta_{u_i} = 0$ and $\sum_{j \in \mathcal{N}^{(H)}_{u_i}} a_{i,j} \Delta^-_j = 0 \geq \theta_{u_i}$. The node u_c will not deviate from investing since $\sum_{u_i \in \mathcal{N}^{(H)}_{u_c}} a_{u_c,u_i} \Delta^-_{u_i} = a \cdot \sum_{i \in S} v_i \geq a \cdot C = \theta_{u_c}$.

For the converse direction, suppose that there is a subset S of indices i such that the actions/edges (u_c, u_i) for $i \in S$ have total cost at most $B = W$, and such that adding the edges (u_c, u_i) makes $\boldsymbol{x}^*$ a PSNE. Because u_c has no incentive to deviate, $\sum_{i \in S} v_i \geq C$. Thus, the set S solves the KNAPSACK problem.

B Proof of Theorem 5.5

We consider the decision version of ANM with an input parameter B. It is clear that the problem is in NP. To prove hardness, we reduce from the 3-PARTITION problem. In the 3-PARTITION problem, we are given a set $S = \{x_1, \ldots, x_{3m}\}$ of $3m$ numbers x_i, each bounded by a polynomial in m. Writing $s = \sum_i x_i$ for the sum of all numbers, the numbers further satisfy that $\frac{s}{4m} < x_i < \frac{s}{2m}$. The goal is to decide if S can be partitioned into disjoint triples $T_1, \ldots, T_m$ such that each triple [7] has the same sum, i.e., $\sum_{x_k \in T_j} x_k = \frac{1}{m} \cdot \sum_{k=1}^{3m} x_k$. The 3-PARTITION problem is strongly NP-complete [Garey and Johnson, 1979], meaning that it is NP-complete even when the x_i are polynomially bounded in m.[8]

Given an instance of the 3-PARTITION problem, we construct an instance of the ANM problem as follows. Let ϵ be a small (though polynomial) constant, and $a > 0$ an arbitrary constant. The set V of agents comprises two types: a set V_1 of $3m$ agents u_i (each corresponding to one number x_i) and a set V_2 of m agents v_j (each corresponding to one triple T_j). The target investment profile $\boldsymbol{x}^*$ is that all agents invest. The graph H is a clique on V. The input altruism network G^{in} is a complete graph on V_1.

For each node $u_i \in V_1$, let $\Delta^-_{u_i} = x_i$, and $\theta_{u_i} = a \cdot (s + \epsilon - x_i)$. For each node $v_j \in V_2$, let $\Delta^-_{v_j} = \epsilon$ and $\theta_{v_j} = a \cdot s/m$. This setup can be realized by, e.g., setting g_{u_i} and g_{v_j} to be linear functions with slopes x_i and ϵ, respectively, w.r.t. the second argument. In addition, the g_{u_i} and g_{v_j} are constant functions w.r.t. the first argument and $c_{u_i} = a \cdot (s + \epsilon - x_i)$ and $c_{v_j} = a \cdot s/m$, which ensures that we have the above θ_{u_i} and θ_{v_j}. The addition/deletion cost for each edge is 1, and the total budget for additions/deletions is $3m$. The reduction obviously takes polynomial time.

First, suppose that there are triples $T_1, \ldots, T_m$ such that $\sum_{x_k \in T_j} x_k = s/m$ for all j. Consider adding the edges (u_i, v_j) if and only if $x_i \in T_j$. This adds exactly $3m$ edges, so it satisfies the budget constraint. Each node $u_i \in V_1$ is incident on exactly one new edge to a node v_j with $\Delta^-_{v_j} = \epsilon$. u_i will not deviate from investing as in the resulting altruism network G (where recall that all off-diagonal entries equal the constant a),

$$\sum_{j \in V \setminus \{u_i\}} a_{u_i,j} \Delta^-_j = a(\epsilon + \sum_{j \neq i} x_j) \ = \ \theta_{u_i}.$$

No node v_j will deviate from investing since

$$\sum_{i \in V \setminus \{v_j\}} a_{j,i} \Delta^-_i = \sum_{x_i \in T_j} a \cdot \Delta^-_i \ = \ a \cdot s/m \ = \ \theta_{v_j}.$$

Next, we show the converse direction. Suppose that there is a set of at most $3m$ edges E, such that the modified altruism network is produced by adding E to A^{in} and that all agents investing is a PSNE of the resulting game. As $\theta_{u_i} > a \cdot (s - x_i)$ for all i, each agent u_i must be incident on at least one edge to some v_j. Because there are $3m$ edges in total, each node u_i must be incident on *exactly* one edge to some v_j. Thus, we can define a partition $S_1, \ldots, S_m$ of $\{x_1, \ldots, x_{3m}\}$ via $S_j = \{x_i \mid u_i \text{ and } v_j \text{ are connected}\}$. For each j, because v_j invests in $\boldsymbol{x}^*$ (which is a PSNE), we must have that

[7] The restriction on the x_i ensures that if there is *any* way to partition the numbers into m sets of equal sum, each set must be a triple.

[8] Recall that the KNAPSACK problem we reduced from earlier is NP-hard only when the values and weights can be exponentially large in the number of items.

$a \cdot \sum_{x_i \in S_j} x_i \geq a \cdot s/m$. But because $\sum_j \sum_{x_i \in S_j} x_i = s$ (because the S_j form a partition), all of the preceding inequalities hold with equality. Thus, the S_j form a partition into sets of the same sum, and by the restriction on the x_i values, they must be a partition into triples.

C Proof of Theorem 5.6

Given an arbitrary action profile $\boldsymbol{x}^*$, checking the feasibility of ANM is equivalent to deciding if the principal can modify the original altruism network G^{in} (by adding/removing edges), such that $\boldsymbol{x}^*$ is an equilibrium of BNPG(G).

We now show NP-hardness even when the principal has an infinite budget, or — equivalently — when all edge costs $\gamma^{(k)}$ are 0. We prove NP-hardness again by reducing from the 3-PARTITION problem, and use the same notation as in the proof of Theorem 5.5.

In the 3-PARTITION problem, we are given a set $S = \{x_1, \ldots, x_{3m}\}$ of $3m$ numbers x_i, each bounded by a polynomial in m. Writing $s = \sum_i x_i$ for the sum of all numbers, the numbers further satisfy that $\frac{s}{4m} \leq x_i \leq \frac{s}{2m}$. The goal is to decide if S can be partitioned into disjoint triples $T_1, \ldots, T_m$ such that each triple[9] has the same sum, i.e., $\sum_{x_k \in T_j} x_k = \frac{1}{m} \cdot \sum_{k=1}^{3m} x_k$.

Given an instance of the 3-PARTITION problem, we construct an instance of the ANM problem as follows. The set V of agents again comprises two types: a set V_1 of $3m$ agents u_i (each corresponding to one number x_i) and a set V_2 of m agents v_j (each corresponding to one desired triple T_j). The target investment profile $\boldsymbol{x}^*$ is that the m agents in V_2 invest while the $3m$ agents in V_1 do not invest. The graph H is again a complete graph on $V_1 \cup V_2$. The input altruism network G^{in} is an empty graph, and, as stated above, all edge additions cost 0.

For each node $u_i \in V_1$, let $\Delta^-_{u_i} = \Delta^+_{u_i} = x_i$, and $\theta_{u_i} = a$. For each node $v_j \in V_2$, let $\Delta^-_{v_j} = \Delta^+_{v_j} = 1$ and $\theta_{v_j} = a \cdot (s/m + m - 1)$. This setup can be realized by, e.g., setting g_{u_i} and g_{v_j} to be linear functions with slopes x_i and 1, respectively, w.r.t. the second argument. In addition, the g_{u_i} and g_{v_j} are constant functions w.r.t. the first argument and $c_{u_i} = a$ and $c_{v_j} = a \cdot (s/m + m - 1)$, which ensures that we have the above θ_{u_i} and θ_{v_j}. As mentioned previously, the addition/deletion cost for each edge is 0; equivalently, the principal's budget is infinite. The reduction obviously takes polynomial time.

First, suppose that there are triples $T_1, \ldots, T_m$ such that $\sum_{x_k \in T_j} x_k = s/m$ for all j. Consider adding the edges (u_i, v_j) if and only if $x_i \in T_j$, and also adding the edges $(v_j, v_{j'})$ for all $j \neq j'$. Each node $u_i \in V_1$is incident on exactly one new edge to a node v_j with $\Delta^-_{v_j} = 1$. Not investing is still a best response for u_i, because

$$\sum_{j \in V \setminus \{u_i\}} a_{u_i,j} \Delta^+_j = a \cdot 1 \; = \; \theta_{u_i};$$

where recall that all off-diagonal entries equal the constant a.

Also, no node v_j will deviate from investing since

$$\begin{aligned} \sum_{i \in V \setminus \{v_j\}} a_{j,i} \Delta^-_i &= \sum_{x_i \in T_j} a \cdot \Delta^-_i + a \cdot (m-1) \\ &= a \cdot (s/m + m - 1) \\ &= \theta_{v_j}. \end{aligned}$$

Next, we show the converse direction. Suppose that there is a set of edges E, such that under the altruism network A^{in} with edge set E, the profile $\boldsymbol{x}^*$ is a PSNE of the game. Because $\Delta^+_{u_i}, \Delta^+_{v_j} \geq 1$ for all nodes u_i and v_j, and $\theta_{u_i} = a$, each node u_i can be incident on at most one edge in E; otherwise, u_i would deviate to investing.

Now consider a node v_j, which must be investing under $\boldsymbol{x}^*$. Therefore,

$$\sum_{i \in V \setminus \{v_j\}} a_{j,i} \Delta^-_i \; \geq \; \theta_{v_j}.$$

By uniform edge weights and the definitions of the Δ^-_i, the sum is equal to $a \cdot (\sum_{u_i:(u_i,v_j) \in E} x_i + \sum_{v_{j'}:(v_{j'},v_j) \in E} 1)$, and by definition, $\theta_{v_j} = a \cdot (s/m + m - 1)$. Therefore, we obtain that $\sum_{u_i:(u_i,v_j) \in E} x_i + \sum_{v_{j'}:(v_{j'},v_j) \in E} 1 \geq s/m + m - 1$. Because the number of neighbors of v_j in V_2 is at most $m - 1$, this implies in particular that $\sum_{u_i:(u_i,v_j) \in E} x_i \geq s/m$. Summing this inequality over all nodes $v_j \in V_2$, we get that $\sum_j \sum_{u_i:(u_i,v_j) \in E} x_i \geq s$. But because each node u_i occurs in the sum for at most one node v_j, each x_i occurs at most once, so we also get that $\sum_j \sum_{u_i:(u_i,v_j) \in E} x_i \leq s$, which implies equality, and thus also that $\sum_{u_i:(u_i,v_j) \in E} x_i = s/m$ for all j. (If the inequality were strict for some j, then it would have to be violated for some other j' for the sum to equal s.)

Thus, the sets $S_1, \ldots, S_m$ defined via $S_j = \{x_i \mid u_i \text{ and } v_j \text{ are connected}\}$ form a partition of $\{x_1, \ldots, x_{3m}\}$, such that $\sum_{x_i \in S_j} x_i = s/m$. Furthermore, by the restriction on the x_i values, they must be a partition into triples.

D Proof of Theorem 5.7

We show the theorem through a non-trivial reduction to the MIN-COST PERFECT MATCHING problem via a connection to another related problem: NETWORK DESIGN FOR DEGREE SETS (NDDS) [Kempe *et al.*, 2020]. We begin by introducing the NDDS problem.

Definition D.1 (NDDS)**.** *Given a simple, undirected, and loop-free graph $\hat{G}^{in} = (V, \hat{E}^{in})$, costs $\gamma^{(i,j)} \geq 0$ for each node pair*[10] *(i, j), and target degree sets D_i which are intervals, i.e., of the form $\{\ell_i, \ell_i + 1, \ldots, r_i\}$. The problem NDDS is to change $\hat{E}^{in}$ to a graph $\hat{E} = (V, \hat{E})$ by adding/deleting edges, such that the total cost is minimized and for all nodes i, the resulting degree $d_{\hat{G}}(i) \in D_i$.*

Kempe *et al.* [2020] showed that the NDDS problem as defined in Definition D.1 can be solved in polynomial time. The formal statement is below:

[9]The restriction on the x_i ensures that if there is *any* way to partition the numbers into m sets of equal sum, each set must be a triple.

[10]If $(i, j) \in \hat{E}^{\text{in}}$, then this is the cost of removing (i, j); otherwise, it is the cost of adding (i, j).

Lemma D.1 (Theorem 4.1 of Kempe *et al.* [2020])**.** *The NDDS problem can be solved in polynomial time, via a reduction to weighted non-bipartite minimum-cost matching.*

When the agents have USL utility functions, the marginal benefits are all identical, i.e., $\Delta_i^- = \Delta_i^+ = \Delta$. In the following discussion, we assume that Δ is strictly positive. Otherwise, there is no motivation for network changes: whether $\boldsymbol{x}^*$ is a PSNE of $\mathrm{BNPG}(G)$ is solely determined by $\theta_1, \ldots, \theta_n$.

The intuition for the reduction is as follows: for any pair of nodes (i, j) which do not share an edge in H, an edge in the altruism network is useless, so we make the cost of such edges (i, j) infinite. When i and j do affect each other, an edge between them contributes Δ to the total altruistic utility experienced by i. Thus, whether i crosses her investment threshold or not only depends on the *number* of neighbors she has in the altruism network G, i.e., on her degree. When i is to invest, this number must be large enough (at least a given threshold), while if i is not to invest, it must be small enough (below a given threshold).

We now describe the reduction more formally. Slightly abusing notation, a BNPG with underlying altruism network G is denoted by $\mathrm{BNPG}(G)$. The input is an instance of ANM, consisting of an input altruism graph $G^{\text{in}} = (V, E^{\text{in}})$, the desired PSNE $\boldsymbol{x}^*$, and costs $\gamma^{(i,j)}$ for edge addition/deletion. In what follows, we show how to construct an instance of NDDS on a graph $\hat{G}^{\text{in}} = (V, \hat{E}^{\text{in}})$ and show that the costs of solving the two instances are equal. As suggested before, $\hat{E}^{\text{in}}$ contains the (undirected) edge (i, j) if and only if $j \in \mathcal{N}_i^{(H)} \cap \mathcal{N}_i^{(G^{\text{in}})}$; in other words, edges such that i's actions do not affect j are removed from the altruism network. Next, we define the target degree sets D_i for each node. When $i \in \mathcal{I}$, we define $D_i := \{\lceil \theta_i/(a \cdot \Delta) \rceil, \ldots, n-1\}$; when $i \notin \mathcal{I}$, the set is $D_i := \{0, \ldots, \lfloor \theta_i/(a \cdot \Delta) \rfloor\}$. Finally, we define the cost $\hat{\gamma}^{(i,j)}$ associated with a node pair (i, j). When $(i, j) \in \hat{E}^{\text{in}}$, we set $\hat{\gamma}^{(i,j)} = \gamma^{(i,j)}$. When $(i, j) \notin \hat{E}^{\text{in}}$, if $i \in \mathcal{N}_j^{(H)}$, the cost is $\hat{\gamma}^{(i,j)} = \gamma^{(i,j)}$; otherwise, it is $\hat{\gamma}^{(i,j)} = \infty$, meaning that the edge (i, j) can not be added. It is clear that the construction runs in polynomial time. The correctness of the construction is captured by Theorem D.2, By Lemma D.1, NDDS is polynomial-time solvable; therefore, we obtain a polynomial-time algorithm for this variant of ANM.

Theorem D.2. *The minimum cost of modifying $G^{in} = (V, E^{in})$ to a graph $G = (V, E)$ such that $\boldsymbol{x}^*$ is a PSNE of* $\mathrm{BNPG}(G)$ *is equal to the minimum cost of modifying $\hat{G}^{in} = (V, \hat{E}^{in})$ to a graph $\hat{G} = (V, \hat{E})$ such that $d_{\hat{G}}(i) \in D_i$ for all i.*

Proof. Because $\Delta_i^- = \Delta_i^+ = \Delta$ for all i and n_i, the action profile $\boldsymbol{x}^*$ is a PSNE of $\mathrm{BNPG}(G)$ if and only if:

$$\begin{aligned} |\mathcal{N}_i^{(H)} \cap \mathcal{N}_i^{(G)}| &\geq \theta_i/(a \cdot \Delta) \quad \text{for all } i \in \mathcal{I} \\ |\mathcal{N}_i^{(H)} \cap \mathcal{N}_i^{(G)}| &\leq \theta_i/(a \cdot \Delta) \quad \text{for all } i \notin \mathcal{I}. \end{aligned} \tag{6}$$

For the first direction, let $G = (V, E)$ be an undirected unweighted altruism network such that $\boldsymbol{x}^*$ is a PSNE of $\mathrm{BNPG}(G)$. Let $B = \sum_{e \in E^{\text{in}} \triangle E} \gamma_e$ be the cost of producing G. We construct a graph $\hat{G}$ such that $d_{\hat{G}}(i) \in D_i$ for all i.

Given the graphs H and G, the construction of $\hat{G} = (V, \hat{E})$ is similar to that of $\hat{G}^{\text{in}}$. $\hat{G}$ contains the edge (i, j) iff $j \in \mathcal{N}_i^{(H)} \cap \mathcal{N}_i^{(G)}$. The neighborhood of a node i is thus $\mathcal{N}_i^{(\hat{G})} = \{j \mid j \in \mathcal{N}_i^{(H)} \cap \mathcal{N}_i^{(G)}\}$. The degree of the node is $d_{\hat{G}}(i) = |\mathcal{N}_i^{(H)} \cap \mathcal{N}_i^{(G)}|$. As $\boldsymbol{x}^*$ is a PSNE of $\mathrm{BNPG}(G)$, Equation (6) implies that for a player $i \in \mathcal{I}$ (respectively, $i \notin \mathcal{I}$), the degree satisfies $d_{\hat{G}}(i) \geq \lceil \theta_i/(a \cdot \Delta) \rceil$ (resp. $d_{\hat{G}}(i) \leq \lfloor \theta_i/(a \cdot \Delta) \rfloor$). Thus, the degree $d_{\hat{G}}(i) \in D_i$ for all i.

Now, we show that the cost of modifying $\hat{G}^{\text{in}}$ to $\hat{G}$ is equal to B. Let $\Gamma := E \triangle E^{\text{in}}$ be the modification from G^{in} to G. Similarly, the modification from $\hat{G}^{\text{in}}$ to $\hat{G}$ is $\hat{\Gamma} := \hat{E} \triangle \hat{E}^{\text{in}}$. We now show that $\Gamma = \hat{\Gamma}$, in particular implying that the proposed modification has cost at most B.

First, observe that every edge $(i, j) \in \Gamma$ must be in H. Otherwise, adding/deleting (i, j) has no impact on altruistic behavior, so B would not have minimum cost. We now reason as follows:

- By definition, $(i, j) \in E \setminus E^{\text{in}}$ iff i and j are neighbors in H and G, but not in G^{in}. This is the case if and only if (i, j) are not neighbors in $\hat{G}^{\text{in}}$, but are neighbors in $\hat{G}$, which is the case if and only if $(i, j) \in \hat{E} \setminus \hat{E}^{\text{in}}$.
- Similarly, $(i, j) \in E^{\text{in}} \setminus E$ iff i and j are neighbors in H and G^{in}, but not in G. This is the case if and only if $(i, j) \in \hat{E}^{\text{in}} \setminus \hat{E}$

Therefore, we have shown that $\Gamma = \hat{\Gamma}$

For the converse direction, we let $\hat{G}$ be the minimum-cost modification of $\hat{G}^{\text{in}}$ such that $d_{\hat{G}}(i) \in D_i$ for all i. Let B be the cost of modifying $\hat{G}^{\text{in}}$ to $\hat{G}$. We first observe that every edge $(i, j) \in \hat{\Gamma}$ must be in H. If $(i, j) \in \hat{E}^{\text{in}} \setminus \hat{E}$, this follows from the definition of $\hat{E}^{\text{in}}$, whereas if $(i, j) \in \hat{E} \setminus \hat{E}^{\text{in}}$, it is because the cost of adding (i, j) would be infinite if (i, j) were not in H.

Next, we construct the graph G. By the construction of $\hat{G}^{\text{in}}$, $(i, j) \in \hat{E}^{\text{in}} \setminus \hat{E}$ implies that i and j are connected in G^{in}, whereas $(i, j) \in \hat{E} \setminus \hat{E}^{\text{in}}$ implies that i and j are *not* connected in G^{in}. In the former case, (i, j) is deleted from G^{in}, i.e., $(i, j) \in E^{\text{in}} \setminus E$, while in the latter case, (i, j) is added to G^{in}, i.e., $(i, j) \in E \setminus E^{\text{in}}$. This modification is applied to all edges in $\hat{\Gamma}$, producing a graph G at cost B.

It remains to show that $\boldsymbol{x}^*$ is a PSNE of $\mathrm{BNPG}(G)$. Consider any node $i \in V$. By construction of G, the set of neighbors added for i is $\mathcal{N}_i^{(H)} \cap (\mathcal{N}_i^{(G)} \setminus \mathcal{N}_i^{(G^{\text{in}})})$. Similarly, the set of neighbors removed from i is $\mathcal{N}_i^{(H)} \cap (\mathcal{N}_i^{(G^{\text{in}})} \setminus \mathcal{N}_i^{(G)})$.

The degree of node i in the graph $\hat{G}$ can be expressed in these terms as

$$
\begin{aligned}
d_{\hat{G}}(i) &= d_{\hat{G}^{\text{in}}}(i) + |\mathcal{N}_i^{(H)} \cap (\mathcal{N}_i^{(G)} \setminus \mathcal{N}_i^{(G^{\text{in}})})| \\
&\quad - |\mathcal{N}_i^{(H)} \cap (\mathcal{N}_i^{(G^{\text{in}})} \setminus \mathcal{N}_i^{(G)})| \\
&= |\mathcal{N}_i^{(H)} \cap \mathcal{N}_i^{(G^{\text{in}})}| + |\mathcal{N}_i^{(H)} \cap (\mathcal{N}_i^{(G)} \setminus \mathcal{N}_i^{(G^{\text{in}})})| \\
&\quad - |\mathcal{N}_i^{(H)} \cap (\mathcal{N}_i^{(G^{\text{in}})} \setminus \mathcal{N}_i^{(G)})| \\
&= |\mathcal{N}_i^{(H)} \cap \mathcal{N}_i^{(G)}|.
\end{aligned}
$$

As $d_{\hat{G}}(i) \in D_i$, it follows that $|\mathcal{N}_i^{(H)} \cap \mathcal{N}_i^{(G)}| \geq \lceil \theta_i/(a \cdot \Delta) \rceil$ for $i \in \mathcal{I}$, and $|\mathcal{N}_i^{(H)} \cap \mathcal{N}_i^{(G)}| \leq \lfloor \theta_i/(a \cdot \Delta) \rfloor$ for $i \notin \mathcal{I}$. Thus, $\boldsymbol{x}^*$ is a PSNE of $\mathrm{BNPG}(G)$. $\square$