# COHESION AND COORDINATION FOR STRATEGIC COLLECTIVE ACTION IN CONTENT RECOMMENDATION

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Increasingly, algorithmic systems are mediating which content users are recommended, thus shaping their exposure to creative media. As awareness of this curation process increases, users have begun to organize deliberate, strategic responses in an attempt to influence which content is promoted online. Existing theoretical models of this behavior, spanning both adversarial "attack" and, more recently, algorithmic collective action literature often represent participating users as having no prior histories and adopting a single unified strategy, effectively behaving as a squadron of clones. We argue that this abstraction does not account for two properties central to real human collectives: *cohesion*, the similarity of members before collective action begins, and *coordination*, the extent to which members align their behavior during strategic action. We introduce a novel model of collective action that parameterizes cohesion and coordination as continuous, independently controllable quantities. We derive conditions characterizing when a collective's strategic behavior shifts a target item's representation, recovering the traditional unified setting as a special case while extending the analysis to partially cohesive and partially coordinated collectives. Controlled synthetic experiments validate these findings and show that successful algorithmic collective action does not require users to behave as clones and, more broadly, that the internal structure of a collective can fundamentally shape its ability to influence algorithmic curation.

## 1 INTRODUCTION

Algorithmic systems increasingly mediate which songs, news, and other cultural content people encounter, shaping cultural production and transmission in ways that were previously governed by human curators and institutions (Seaver, 2022). As users have grown aware of this mediating role, many have begun to act on it: developing informal theories about how their behavior — what they stream, rate, or share — feeds back into what a system subsequently recommends, and acting strategically on those theories (Xiao et al., 2025; Davidjants et al., 2026; Karizat et al., 2021). For example, fan communities organizing streaming campaigns to boost a new release are a concrete example of this behavior in practice (Behnke, 2022).

Strategic user influence over recommender systems has historically been studied as a security problem, in which organized accounts are modeled as adversaries manipulating a target item's exposure (Shams & Leith, 2024). A separate, more recent literature has begun to study algorithmic collective action where the coordinating agents are real, value-driven platform users rather than fabricated accounts (Hardt et al., 2023; Fedorova et al., 2026; Baumann & Mendler-Dünner, 2024).

Formally, both lines of work represent the acting group in the same way: a fraction $\alpha$ of the population adopts a single strategy applied uniformly across every member (Hardt et al., 2023; Gauthier et al., 2025; Fedorova et al., 2026; Baumann & Mendler-Dünner, 2024). This abstraction is natural when the acting accounts are created specifically for an attack and have no meaningful prior state. It is much less natural for a collective of existing human users. Before a strategic action begins, those users already have distinct interaction histories, preferences, and learned representations, which impact their ability to engage in unified behaviors. Two recent papers begin to relax this assumption. Karan et al. (2025b) argue that a collective's impact should not be reduced to its size alone, pointing to Ostrom's classical observation that a group's homogeneity, not just its scale, governs its capacity for collective action (Ostrom, 1990), but they do not formalize this claim. Then, Karan et al. (2025a)



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

go further and formally model imperfect coordination as noise around a shared target strategy, deriving bounds showing that sufficiently high coordination noise can reduce a collective's success rate from near-certain to only marginally above baseline. However, their formulation does not represent the members' pre-existing, heterogeneous states, which is not something real collectives can control.

We argue that a model of human collective action needs to separately represent these two properties: *cohesion*, the similarity of a collective's members before any organized strategic action, and *coordination*, the extent to which members subsequently align their behavior toward a shared strategy. We introduce a representation-space model of strategic collective action that separates these two dimensions and models the success of a collective with respect to a learned representational space and their ability to modify that space's geometry via strategic action.

We study the effects of these parameters theoretically and empirically. We first derive the target-item representation shift induced by a heterogeneous, partially coordinated collective and recover the conventional unified-clone model as a special case. We then characterize when the familiar monotonicity and diminishing-returns properties of the clone setting persist, and show that residual heterogeneity can cause both to fail. We use controlled synthetic and real-world experiments to examine how collective size, initial heterogeneity, coordination, and strategy interact under realizable interventions, showing that a collective need not behave in a fully unified fashion to achieve success.

## 2 Related Work

Broadly, collective action describes individuals organizing around a shared objective (Olson, 1965). In algorithmic systems, groups may pursue such objectives by strategically modifying the data they contribute, an idea formalized through notions such as *data leverage* (Vincent et al., 2021). Historically, this behavior is closely related to data poisoning and shilling attacks, which manipulate training data to alter predictions for targeted items (Shams & Leith, 2024; Li et al., 2016; Sundar et al., 2020). In particular, our mathematical model is inspired by Shams & Leith (2024). However, rather than injecting artificial users, we consider existing users strategically modifying their own behavior towards a collective goal and we concretely characterize the roles of proportion, cohesion, and coordination in our theoretical contributions.

Recent work has begun to formally distinguish the setting of data poisoning from *algorithmic collective action*, studying how value-oriented groups of users can coordinate their data contributions to influence learned systems (Hardt et al., 2023; Ben-Dov et al., 2024; Gauthier et al., 2025; Karan et al., 2025a;b), with some specifically considering the setting of recommender systems (Baumann & Mendler-Dünner, 2024; Fang et al., 2022). Baumann & Mendler-Dünner (2024) empirically show that fully coordinated behavior can increase the recommendation of a target item.

We distinguish our work along two concrete axes. First, a central distinction of our framework is that we are able to separately model two properties of a collective that are otherwise difficult to measure within a representational space: *cohesion*, which captures the similarity among collective members prior to strategic action, and *coordination*, which captures the extent to which members subsequently align toward a shared strategy. While recent work has begun to relax assumptions of homogeneous collective behavior, including through heterogeneous collectives and imperfect coordination (Karan et al., 2025b;a), our framework explicitly separates a collective's pre-existing heterogeneity from its subsequent degree of strategic alignment. Separating these properties allows us to characterize their effects on success in collective action. Second, while previous work studying collective action in recommender systems has focused primarily on understanding collective action with respect to ranking outcomes (e.g., changes in the Top-K most relevant items served to a user) (Baumann & Mendler-Dünner, 2024; Karan et al., 2025b; Karizat et al., 2021), our goal is to directly consider the representational shifts that can be induced by collective action.

## 3 Preliminaries

### 3.1 Representation-Based Recommendation Systems

We consider representation-based recommender systems. Let $\mathcal{U}$ denote the set of users and $\mathcal{I}$ the set of items. In such a setting, the goal of the recommender is to learn representations of users and items

$u_i \in \mathbb{R}^d \; \forall i \in \mathcal{U}$ and $v_j \in \mathbb{R}^d \; \forall j \in \mathcal{I}$ for some $d \in \mathbb{N}$. Each of the representations is a function of observed interaction histories:

$$u_i = f_\Theta(\Omega_{i,\mathcal{I}}, x_i), \qquad v_j = g_\Theta(\Omega_{\mathcal{U},j}, x_j),$$

where $\Omega_{i,\mathcal{I}}$ and $\Omega_{\mathcal{U},j}$ denote the interactions associated with user $i$ and item $j$ respectively, $x_i, x_j$ denote any additional user or item-side information, and each model $f_\Theta, g_\Theta$ may be independently trained under a set of parameters $\Theta$. Given learned user and item representations, the recommender assigns a score to each user–item pair through a scoring function

$$\hat{r}_{ij} = s_\Theta(u_i, v_j),$$

where larger values of $\hat{r}_{ij}$ indicate greater predicted relevance of item $j$ to user $i$.

Ordinarily, users will engage in actions according to their real preferences. However, a strategic user may instead choose interactions to affect their future learned representation or global recommendations. In this work, we focus on strategic users whose goal is to promote a specific item $j^\star$ (Baumann & Mendler-Dünner, 2024; Shams & Leith, 2024) (defined formally in Definition 2, below).

Note that, under this formulation, while the underlying representation functions $f_\Theta$ and $g_\Theta$ remain fixed, users can influence the representations assigned to them and various target items by changing their behavioral histories supplied to the model. Thus, strategic behavior can affect recommendation outcomes without directly modifying the recommender itself. In this work, we will use Alternating Least Squares (ALS) (Koren et al., 2009) as an analytically tractable instantiation of the broader representation-based lens through which we study strategic action. In particular, the ALS model allows us to concretely decompose the effects of collective action and analyze the characteristics that enable collectives to manipulate the output of recommender systems.

**ALS Recommendation Model** In ALS, solutions to $\min_{U,V} \sum_{(i,j)\in\Omega_{\mathcal{U},\mathcal{I}}} (R_{ij} - u_i^\top v_j)^2 + \lambda\|U\|_F^2 + \lambda\|V\|_F^2$ yield user and item representations, $U, V$, and predicted ratings are inner products of user/item representations, $s_\Theta(u_i, v_j) = u_i^\top v_j$. Generally, $U, V$ are learned by iteratively solving for one and fixing the other via the closed-form updates:

$$u_i = f_\Theta(\Omega_{i,\mathcal{I}}) = \left( \sum_{j\in\Omega_{i,\mathcal{I}}} v_j v_j^\top + \lambda I \right)^{-1} \left( \sum_{j\in\Omega_{i,\mathcal{I}}} r_{ij} v_j \right),$$

$$v_j = g_\Theta(\Omega_{\mathcal{U},j}) = \left( \sum_{i\in\Omega_{\mathcal{U},j}} u_i u_i^\top + \lambda I \right)^{-1} \left( \sum_{i\in\Omega_{\mathcal{U},j}} r_{ij} u_i \right).$$

# 4 Theoretical Analysis

## 4.1 Modeling Strategic Collective Action

Let $C \subseteq \mathcal{U}$ denote a collective comprising proportion $\alpha = \frac{|C|}{n}$ of the user population. We consider a warm-start setting (Gusak et al., 2025), in which each member $i \in C$ begins with a prior interaction history $\Omega_{i,\mathcal{I}}$ and corresponding representation $u_i = f_\Theta(\Omega_{i,\mathcal{I}}, x_i)$. We let $u_i$ and $u_i'$ denote the representations immediately before and after strategic action, respectively.

**Pre-Action Collective State:** The collective's prior state is characterized by mean and covariance

$$\bar{u}_C = \frac{1}{|C|} \sum_{i\in C} u_i, \qquad \Sigma_C = \frac{1}{|C|} \sum_{i\in C} (u_i - \bar{u}_C)(u_i - \bar{u}_C)^\top. \tag{1}$$

The covariance $\Sigma_C$ captures the collective's initial covariance; a more cohesive collective has members concentrated around $\bar{u}_C$, whereas a less cohesive collective occupies a more dispersed region of the representation space.

**Collective's Rating Strategy:** Let $h \in \{(\mathcal{I}, \mathbb{R})_j\}_{j\in[m]}$ denote a rating strategy, defined by a set of strategic items and their associated strategic ratings, that induces a shared strategic direction $u_h = f_\Theta(h)$. For an example of $f$, see the ALS instantiation in Section 3.1.

Using each of these elements (visualized in Figure 1), we now define:

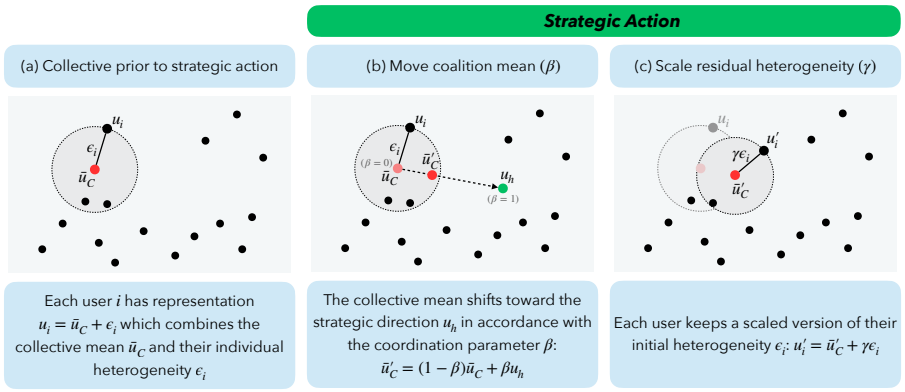



Figure 1: Strategic collective action in representation space. (a) Collective members begin from heterogeneous representations around the prior mean $\bar{u}_C$. (b) The coordination parameter $\beta$ shifts the collective mean toward the strategic direction $u_h$. (c) The parameter $\gamma$ controls how much of the collective's initial heterogeneity around $\bar{u}'_C$ remains after strategic action.

**Definition 1 (Strategic Action).** *Let $\epsilon_i = u_i - \bar{u}_C$ denote member $i$'s deviation from the collective's pre-action mean, where $\frac{1}{|C|}\sum_{i\in C}\epsilon_i = 0$. Given strategic direction $u_h$, we characterize the post-action representation of member $i$ as*

$$u'_i = (1-\beta)\bar{u}_C + \beta u_h + \gamma\epsilon_i, \tag{2}$$

*where $\epsilon_i\epsilon_i^T$ controls the initial* cohesion*, or similarity of a collective's members before any organized strategic action, $\beta \in [0,1]$ controls* coordination*, movement of the collective mean toward $u_h$, and $\gamma \in [0,1]$ controls the retention of members' initial* cohesion*.*

**Post-Action Collective State:** Averaging Equation (2) over collective members gives

$$\bar{u}'_C = (1-\beta)\bar{u}_C + \beta u_h, \qquad \Sigma'_C = \gamma^2\Sigma_C. \tag{3}$$

For a formal derivation of these properties, please see Section D.1.1.

### 4.2 Representational Shifts: Defining the Collective's Goal

We can formalize the collective's objective. Let $j^\star \in \mathcal{I}$ denote a target item that the collective seeks to widely promote. Specifically, the collective seeks to induce a shift in the learned representation of $j^\star$, $v_{j^\star}$ such that, relative to a fixed set of user representations, the item's predicted relevance increases.

**Definition 2.** *[Average Predicted Relevance] Given user representations $U = \{u_i\}_{i\in\mathcal{U}}$ and a target-item representation $v_{j^\star}$, we define the average predicted relevance of item $j^\star$ as*

$$E_{j^\star}(U,V) = \frac{1}{|\mathcal{U}|}\sum_{i\in\mathcal{U}} s_\Theta(u_i, v_{j^\star}) = \frac{1}{|\mathcal{U}|}\sum_{i\in\mathcal{U}} u_i^\top v_{j^\star} = \bar{u}^\top v_{j^\star},$$

*where $s_\Theta$ is the inner-product scoring function of a recommender and $\bar{u} = \frac{1}{|\mathcal{U}|}\sum_{i\in\mathcal{U}} u_i$ denotes the mean user representation.*

**Definition 3.** *[Collective Success: Target Item Shift] Let $(U,V)$ denote the user and item representations before the collective's strategic action $h$, and let $V'$ denote the item representations after a collective of users engages in strategic action (Definition 1). We say that the collective is* successful *if $E_{j^\star}(U,V') > E_{j^\star}(U,V)$. We measure the magnitude of success as $S_{(\alpha,\beta,\gamma,h)} := E_{j^\star}(U,V') - E_{j^\star}(U,V)$, where $\alpha$, $\beta$, $\gamma$, and $h$ parameterize the characteristics of the collective.*

This formulation serves two purposes. First, it provides a characterization of collective action at the level of learned representations. Second, these representation changes have a natural downstream interpretation: increasing an item's predicted relevance for the general user population makes the item more likely to receive higher predicted ratings. Consequently, a promoted target item would receive greater exposure under some top-$k$ recommendation policy (Baumann & Mendler-Dünner, 2024).

### 4.3 CHARACTERIZING SUCCESS

We now characterize how the collective's strategic action changes the representation of the target item $j^\star$ and, consequently, its predicted relevance to the broader population. Let $\Omega_{\mathcal{U}\setminus C,j^\star}$ denote the non-collective users who rated $j^\star$ prior to strategic action. Following the ALS equations in Section 3.1, define

$$A := \sum_{i\in\Omega_{\mathcal{U}\setminus C,j^\star}} u_i u_i^\top + \lambda I, \qquad b := \sum_{i\in\Omega_{\mathcal{U}\setminus C,j^\star}} r_{ij^\star} u_i, \tag{4}$$

such that the pre-action target representation is $v_{j^\star} = A^{-1}b$.

Suppose that collective members follow the strategic-action model in Equation (2) and assign the target item a common rating $r_{Cj^\star} \in \mathbb{R}$. As shown formally in Appendix D.1.2, the collective's strategic action can be seen as a perturbation to the initial matrix $A$, and decomposed into the post-action collective mean $\bar{u}'_C$ and residual heterogeneity $\gamma^2\Sigma_C$. We can define this perturbation as $A' := A + \alpha n\left(\bar{u}'_C\bar{u}'^\top_C + \gamma^2\Sigma_C\right)$, and the resulting effect can be written as

$$v'_{j^\star} - v_{j^\star} = \alpha n (A')^{-1}\left[\underbrace{\bar{u}'_C\left(r_{Cj^\star} - \bar{u}'^\top_C v_{j^\star}\right)}_{\text{coordinated mean signal}} - \underbrace{\gamma^2\Sigma_C v_{j^\star}}_{\text{residual heterogeneity}}\right]. \tag{5}$$

This immediately yields our condition for successful collective action.

**Theorem 1.** *[Collective-Action Success] For a collective characterized by $\alpha$, $\beta$, and $\gamma$ that follows strategy $h$ inducing strategic direction $u_h$, the population-level success is*

$$S_{(\alpha,\beta,\gamma,h)} = \alpha n\, \bar{u}^\top (A')^{-1}\left[\bar{u}'_C\left(r_{Cj^\star} - \bar{u}'^\top_C v_{j^\star}\right) - \gamma^2\Sigma_C v_{j^\star}\right].$$

Theorem 1 exposes two levers of collective influence. The first is a *mean signal*: coordination moves the collective toward $u_h$, producing a signal proportional to the gap between its desired rating $r_{Cj^\star}$ and the target's current predicted relevance $\bar{u}'_C v_{j^\star}$. The second is a *heterogeneity signal*: residual heterogeneity contributes through $-\gamma^2\Sigma_C v_{j^\star}$.

**Unified clones as a special case.** We can also show that our general success is able to recover the conventional unified-clone setting at $\beta = 1, \gamma = 0$. Setting $\beta = 1$ and $\gamma = 0$ gives $\bar{u}'_C = u_h$ and removes the heterogeneity term. The generalized model therefore reduces to the conventional unified strategic-user setting (similar to the formulation of Shams & Leith) where the contribution of the strategic collective is represented as a rank-one perturbation to the matrix $A$). For a more formal proof of this result, please see Appendix D.1.3.

**Corollary 1.** *[Unified-Clone Success] For $\beta = 1$ and $\gamma = 0$,*

$$S_{(\alpha,1,0,h)} = \frac{\alpha n(\bar{u}^\top A^{-1}u_h)(r_{Cj^\star} - u_h^\top v_{j^\star})}{1 + \alpha n u_h^\top A^{-1}u_h}.$$

Next, we consider the dynamics that illustrate the relationship between $\alpha$ and the success achieved by a fully unified clone (i.e., cohesive and coordinated) collective. First, we show the monotonicity of success with respect to $\alpha$ (Proposition 1). Second, we highlight the diminishing returns, where, as the collective grows, each additional member contributes less to the overall success (Proposition 2).

**Proposition 1.** *[Monotonicity of Success Under Unified Clone Setting] Assuming that $\bar{u}^\top A^{-1}u_h(r_{Cj^\star} - u_h^\top v_{j^\star}) > 0$, the collective success $S_{(\alpha,\beta=1,\gamma=0,h)}$ is monotonically increasing in $\alpha \in (0, 1]$.*

**Proposition 2.** *[Diminishing Returns Under Unified Clone Setting] Assuming that $\bar{u}^\top A^{-1}u_h(r_{Cj^\star} - u_h^\top v_{j^\star}) > 0$, $S_{(\alpha,\beta=1,\gamma=0,h)}$ is strictly increasing and strictly concave in $\alpha$. In other words, collective action exhibits diminishing marginal returns in its success with respect to collective size $\alpha$.*

We can now use the more general formulation over success conditions to consider whether the patterns present in the unified clone setting persist in the more realistic collective action setting. We

find that both of these conditions are intrinsically linked with the realized heterogeneity (reflected in the cohesion parameter $\gamma$) of the collective. First, we find that holding $\gamma = 0$ fixed recovers the monotonicity of success and diminishing returns that we observe in the unified clone setting.

**Theorem 2.** *[Monotonicity of Success and Diminishing Returns with $\gamma = 0$] Assuming $(\bar{u}^\top A^{-1}\bar{u}'_C)(r_{Cj^\star} - \bar{u}'^\top_C v_{j^\star}) > 0$, for any constant value $\beta$, if $\gamma = 0$ the collective success $S_{(\alpha,\beta,\gamma=0,h)}$ is monotonically increasing in $\alpha \in (0, 1]$ and is strictly concave in $\alpha$.*

However, when $\gamma > 0$, these properties may not hold. In Appendix D.2, we present counterexamples where increasing collective size $\alpha$ violates both the monotonicity principle and diminishing returns.

## 5 Empirical Evaluation

We organize our results around the two key research questions. We first characterize how collective size, coordination, cohesion, and strategy determine success (RQ1), then we use this to reason about the patterns that emerge only after relaxing the unified-clone assumptions of perfect coordination and cohesion (RQ2).

### 5.1 Experimental Setup

**Recommender Models.** We use ALS to directly validate our theoretical analysis. Architecture and training details are provided in Appendix E.6.

**Evaluation Metrics.** We evaluate collective success at three complementary levels: target-item movement, post-refit relevance, and ranking outcomes. First, our theoretical metric *Item-Shift Success* isolates movement of the target item while evaluating that movement relative to the baseline user population:

$$S_{\text{target item shift}} = \bar{u}^\top \left(v'_{j^\star} - v_{j^\star}\right).$$

This corresponds directly to the success quantity analyzed in Theorem 1. Additionally, in practice, retraining the recommender may alter both user and item representations. We therefore additionally measure the average change in predicted relevance of the target item after refitting ALS on the post-action rating matrix $R'$, *Refitted ALS Success*, as:

$$S_{\text{refit}} = \frac{1}{|\mathcal{U}|} \sum_{i \in \mathcal{U}} \left[(u'_i)^\top v'_{j^\star} - u_i^\top v_{j^\star}\right].$$

This metric captures whether the theoretical target-item shift persists once the full recommender is allowed to adapt to the collective's actions. Finally, we measure whether these representational and score changes translate into changes in recommendation exposure, *Top-10 Change*. Let $T@10(j^\star; \Theta) = \frac{1}{|\mathcal{U}|} \sum_{i \in \mathcal{U}} \mathbb{I}\left[j^\star \in \text{Top}10_i(\Theta)\right]$ denote the fraction of users for whom the target item appears among their top-10 recommendations. We report the pre-to-post action change

$$\text{Change in Top10} = T@10(j^\star; \Theta') - T@10(j^\star; \Theta).$$

**Collective Strategies.** We evaluate four realizable strategies, each assigning the target item $r_{Cj^\star} = 5$ while selecting additional strategic items differently: *Random Filler* samples items uniformly; *Mainstream Mimic* selects globally popular items; *Target Neighborhood* selects items similar to the target; and *Target-Supporter Mimic* selects items preferred by users who already support the target. We additionally compare against an optimal strategy that solves for an optimal direction subject to a movement budget matched to each heuristic. Complete strategy constructions and optimization procedures are given in Appendix E.4. The differences between these strategies can be visualized in Figure 2.

**Synthetic Protocol.** We construct synthetic environments in which we independently control collective size $\alpha$, initial heterogeneity $tr(\Sigma_C)$, coordination $\beta$, heterogeneity retention $\gamma$, and strategy $h$. For each environment, we first train a baseline recommender and obtain pre-action representations $(U, V)$. Our theory specifies strategic action directly in representation space. Thus, for each collective, we first construct the theoretical post-action representations in Definition 1: $u'_i = (1 - \beta)\bar{u}_C + \beta u_h + \gamma\epsilon_i$, then translate these targets into feasible modifications of each member's rating history. These interventions produce a modified ratings matrix $R'$, which we pass

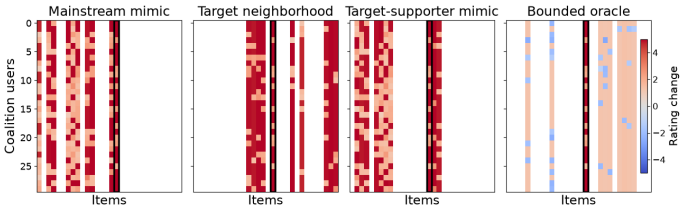



Figure 2: Realized rating modifications $R'$ under the (bounded) optimized and several selected heuristic strategies, shown for $\alpha = 0.3$, $\beta = 1.0$, and $\gamma = 0.0$. Rows correspond to collective members and columns correspond to items. Whereas the realizable heuristics primarily add high ratings to previously unrated strategic items, the optimized strategy additionally reduces ratings assigned to previously rated items. For more extensive visualizations, please see Appendix F.1.

through the frozen baseline encoder to measure the representation-space change actually realized by the intervention. We then refit the recommender on $R'$ and evaluate both the resulting target-item shift and downstream recommendation outcomes. This pipeline separates the *requested* collective state from its *behavioral realization* and from subsequent model adaptation. Full data-generation, strategic action, and training procedures are provided in Appendices E.1 - E.3.

## 5.2 Empirical Contributions

### 5.2.1 RQ1: How do the individual components of our collective action model impact success?

**Representational shift is predictive of longer term recommender system metrics.** We can see from the visualizations in Figure 3 that for all of the heuristic strategies our theoretical success metric of Item Shift Success has similar patterns to that of our more realistic downstream metrics. This correspondence supports our use of target-item movement as an interpretable mechanism for understanding collective influence.

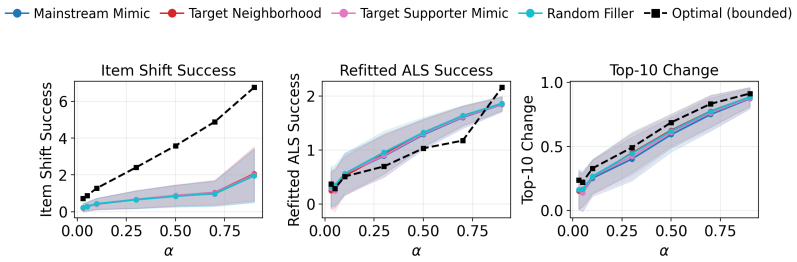
**Collective size amplifies influence.** In Figure 3, we can see that across each of the strategies, increasing the proportion $\alpha$ of users participating in a collective increases the magnitude of the collective's influence on the target item. This effect is visible both in the immediate Item-Shift Success analyzed by our theory and in the downstream Post-Refit Success and Top-10 Change. Thus, as shown in previous literature (Baumann & Mendler-Dünner, 2024), collective size controls the amount of influence available to the collective. We provide additional sweeps of the parameter space over cohesion and coordination value ranges in Appendix F.2, where the patterns exhibited by the fully unified collective persist even under relaxation to a more generalized case.

**Relaxing unified-clone assumptions narrows the gap between optimized and heuristic strategies.** Under the generalized collective-action setting, the advantage of the movement-matched optimal strategy largely disappears. As shown in Figure 3(b), the realizable heuristic strategies closely track the optimized strategy in Item-Shift Success across collective sizes and produce nearly identical changes in Top-10 exposure. After model refitting, several heuristics even outperform the strategy optimized for immediate item shift. This suggests that the large advantage of the latent-space optimum observed under the unified-clone abstraction is not intrinsic to collective action itself. Once participants are permitted to retain heterogeneous representations and coordinate imperfectly, simple feasible behaviors can realize comparable downstream influence.

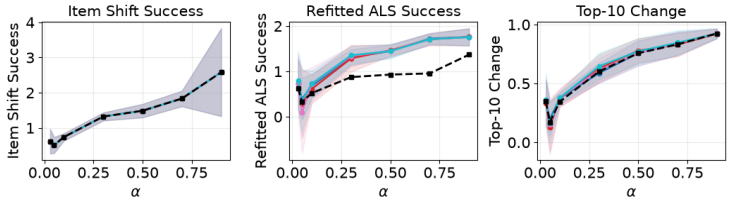
### 5.2.2 RQ2: What does the collective-action model reveal beyond the unified-clone setting?

**Greater initial heterogeneity does not uniformly reduce success.** Figure 3(c) varies the collective's initial dispersion $\mathrm{tr}(\Sigma_C)$ while holding $\beta = 1$ and $\gamma = 1$. The bounded optimal strategy exhibits the expected degradation in immediate Item-Shift Success as the collective becomes more heterogeneous. In contrast, the realizable heuristic strategies are considerably less sensitive to initial dispersion: their Item-Shift curves remain close to flat, while downstream Post-Refit Success and Top-10 Change vary non-monotonically across heterogeneity levels. This indicates that pre-existing heterogeneity is not simply a penalty that uniformly weakens collective influence.

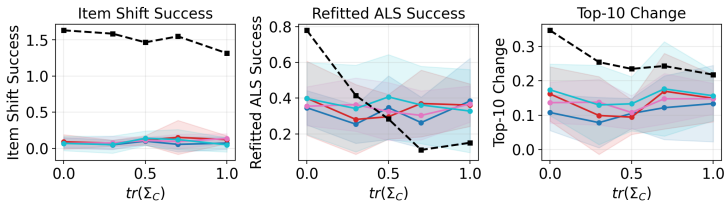
**More coordination can make a realizable collective less successful.** The clearest departure from the unified-clone intuition appears in Figure 3(d). As $\beta$ increases, the bounded optimal strategy ben-



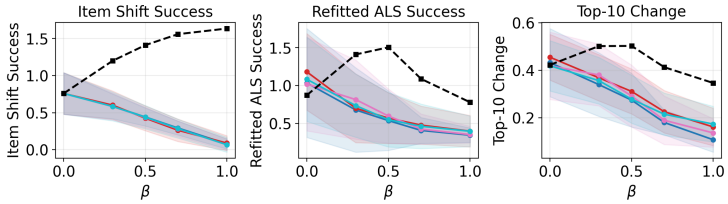

(a) Results for varying collective proportion $\alpha$ under $\beta = 1, \mathrm{tr}(\Sigma_C) = \gamma = 0$



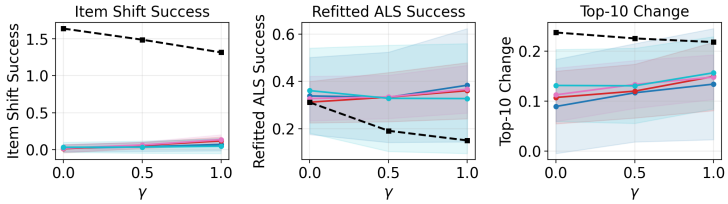

(b) Results for varying collective proportion $\alpha$ under $\beta = 0, \mathrm{tr}(\Sigma_C) = \gamma = 1$




(c) Results for varying cohesion $\mathrm{tr}(\Sigma_C)$ under $\gamma = 1, \beta = 1, \alpha = 0.1$




(d) Results for varying coordination $\beta$ under $tr(\Sigma_C) = \gamma = 0, \alpha = 0.1$




(e) Results for varying heterogeneity retention $\gamma$ under $\mathrm{tr}(\Sigma_C) = 1, \beta = 1, \alpha = 0.1$

Figure 3: Effects of collective size, initial heterogeneity, coordination, and retention of heterogeneity on collective success. Results are averaged over 10 independent seeds. In each row, we vary one component of the collective-action model while fixing the others. Meanwhile, for each column we iterate over the success metrics defined in Section 5.1. For more comprehensive sweeps over all the parameter combinations, please see Appendix F.2.

efits from stronger movement toward its chosen strategic direction. The realizable heuristic strategies show the opposite pattern: increasing coordination generally decreases Item-Shift Success, and the same downward trend appears after refitting and in Top-10 exposure. Thus, coordination is not intrinsically beneficial. Figure 2 helps explain why the optimized and realizable strategies behave

differently. The optimized strategy is free to modify previously observed ratings in ways that directly maximize Item-Shift Success, whereas the heuristic strategies act primarily through feasible additions to users' interaction histories.

**Residual heterogeneity doesn't sacrifice downstream influence.** Figure 3(e) varies $\gamma$, which controls how much of the collective's initial heterogeneity remains after strategic action. For the bounded optimal strategy, increasing retention of heterogeneity reduces refitted and downstream success. In contrast, the realizable heuristic strategies are largely insensitive to $\gamma$ and in several cases show slight improvements in Post-Refit Success and Top-10 Change as more heterogeneity is retained. In other words, successful strategic action does not require the collective to collapse to a single post-action representation.

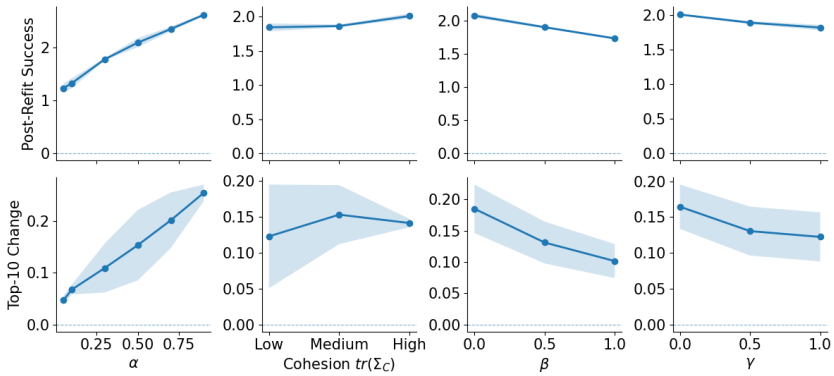



Figure 4: MovieLens-100K validation of the collective action patterns. We report downstream Post-Refit Success and Top-10 Change as collective size, initial heterogeneity, coordination, and retention of heterogeneity vary. We showcase the *Mainstream Mimic* heuristic, averaged across 5 seeds.

**The same qualitative behaviors appear in real interaction data.** We repeat the ALS experiments on MovieLens-100K, replacing exact latent-space control of initial cohesion with low-, medium-, and high-heterogeneity sampling regimes. As shown in Figure 4, the real-data experiment preserves the central qualitative conclusions of the synthetic analysis: (1) larger collectives exert greater downstream influence, (2) perfect coordination is not necessary for success and can reduce downstream success metrics, and (3) perfect cohesion is not required for successful collective action. For more details of our real-world experiments, please see Appendix F.3. Together, these results reveal the key distinction between heterogeneous collective action and the unified-clone setting: Successful collective action does not require participants to collapse onto a single representation.

## 6 Conclusion

We study strategic collective action in recommender systems under a setting where participating users are existing platform users rather than interchangeable strategic clones. Our results support two main conclusions. First, modeling human collective action requires separating a collective's pre-existing heterogeneity from the extent to which its members subsequently coordinate toward a shared strategy. Second, we find that greater coordination and greater cohesion are not uniformly beneficial under realizable strategies. Taken together, our findings suggest that collective influence over recommender systems cannot be understood through collective size alone. The internal structure of the collective—where its members begin, how they coordinate, and how much heterogeneity they retain—is itself part of the mechanism through which collective action shapes algorithmic curation. Future work could extend the analysis to a broader set of recommendation models via influence functions (Pruthi et al., 2020; Zhang et al., 2024). More broadly, these findings suggest that influence over algorithmic curation is not limited to centralized synthetic actors. Groups of real users, acting collectively but imperfectly, can also shape the representations and exposure patterns produced by recommender systems.

## 7 AI DISCLOSURE SECTION

In this work, we used generative AI tools for:

- formulating mathematical claims,
- providing critical ingredients for proving mathematical claims, and
- providing feedback on experiments, implementing methods.

In particular, we used an AI assistant to help review portions of a Matrix Perturbation textbook (Stewart & Sun, 1990) and refine our proof techniques. We have fully reviewed all the theoretical portions and rewritten them ourselves several times to increase the clarity. Additionally, given the complexity of our synthetic experiments and the necessary compatibility with the tensorflow implementations of our two-tower model architectures, we occasionally used an AI coding assistant for writing the boilerplate portions of the code (e.g. dataloading, generating plots, iterating to improve the cleanliness and efficiency of our codebase). All the final code was written by us and tested.

We have not used generative AI tools for

- generating synthetic data sets,
- helping to develop theoretical models or conceptual frameworks,
- proposing or refining hypotheses,
- assisting with translation, or
- interpreting results.

Meanwhile,

- supporting qualitative and thematic data analysis
- and cleaning and reformatting datasets

are not applicable to this work. Additionally, we used generative AI tools for brainstorming and proposing a title for the research paper, and editing to improve readability (specifically, shortening). During the research process, our authors used different AI assistants to reflect on their ideas such that they could be communicated succinctly during group meetings.

We have reviewed all AI-assisted work, many times and tried our best to make sure that all documentation (both code and proofs) was written without AI assistance. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

## 8 ETHICS STATEMENT

This work studies how coordinated groups of users can strategically alter their interactions with recommender systems in order to influence subsequent recommendation outcomes. As such, the mechanisms analyzed in this work are potentially dual-use: similar techniques could be employed for legitimate collective action, such as coordinated efforts by communities or creators to increase the visibility of underrepresented content, but could also be adapted for manipulative behavior such as artificial promotion or coordinated platform gaming.

Our empirical analysis relies on synthetic environments and publicly available recommendation datasets and does not involve interventions on deployed platforms or interactions with real users. We do not conduct experiments that manipulate real-world recommendation systems, platform users, or content exposure. We nevertheless acknowledge that the behavioral strategies and theoretical insights developed here could potentially be repurposed outside the intended research context, and encourage future work to consider safeguards that distinguish legitimate forms of collective participation from deceptive or harmful manipulation.

## 9 Reproducibility Statement

We take several steps to support the reproducibility of our theoretical and empirical results. All assumptions and derivations underlying our theoretical analysis are stated explicitly in the main text, with complete proofs provided in Appendix D. Additionally, we provide detailed descriptions of our synthetic data generation process, collective construction procedure and strategic action strategies in Appendix E. We also detail our model architectures, training procedures and hyperparameters, and evaluation metrics in the experimental sections and supplementary material in Appendix E. Finally, code for reproducing the experiments on both synthetic and real-world data can be accessed at `https://anonymous.4open.science/r/CollectivesNotClones-090D/README.md`.

All experiments were completed on an M1 Macbook Pro. Evaluating a full sweep with $\alpha \in (0.01, 0.03, 0.05, 0.1, 0.3, 0.5, 0.7, 0.9), \beta \in (0.0, 0.3, 0.5, 0.7, 1.0), \gamma \in (0.0, 0.5, 1.0), h \in \{\text{random filler, mainstream mimic, target neighborhood, target supporter}\}$ took approximately 2 hours per seed.

## References


Joachim Baumann and Celestine Mendler-Dünner. Algorithmic collective action in recommender systems: Promoting songs by reordering playlists. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 119123–119149. Curran Associates, Inc., 2024. doi: 10.52202/079017-3784. URL `https://proceedings.neurips.cc/paper_files/paper/2024/file/d79792543133425ff79513c147dc8881-Paper-Conference.pdf`.

Taylor Behnke. The blackpink fans looking to outfox youtube. *Vulture*, September 2022. URL `https://www.vulture.com/2022/09/blackpink-fans-born-pink-youtube-streams.html`. Accessed: 2026-09-15.

Omri Ben-Dov, Jake Fawkes, Samira Samadi, and Amartya Sanyal. The role of learning algorithms in collective action. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 3443–3461. PMLR, 21–27 Jul 2024. URL `https://proceedings.mlr.press/v235/ben-dov24a.html`.

Jaana Davidjants, Josie Hamper, Matti Kortesoja, Katrin Tiidenberg, Andrea Schaffar, and Jenni Niemelä-Nyrhinen. Consumptive curation for self-care: in search of clumsier algorithms. *Information, Communication & Society*, 0(0):1–17, 2026. doi: 10.1080/1369118X.2026.2659282. URL `https://doi.org/10.1080/1369118X.2026.2659282`.

Minghong Fang, Jia Liu, Michinari Momma, and Yi Sun. Fairroad: Achieving fairness for recommender systems with optimized antidote data. In *Proceedings of the 27th ACM on Symposium on Access Control Models and Technologies*, SACMAT '22, pp. 173–184, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450393577. doi: 10.1145/3532105.3535023. URL `https://doi.org/10.1145/3532105.3535023`.

Ekaterina Fedorova, Madeline Kitch, and Chara Podimata. Can users fix algorithms? a game-theoretic analysis of collective content amplification in recommender systems, 2026. URL `https://arxiv.org/abs/2506.04525`.

Etienne Gauthier, Francis Bach, and Michael I. Jordan. Statistical collusion by collectives on learning platforms. In *Forty-second International Conference on Machine Learning*, 2025. URL `https://openreview.net/forum?id=46yLEXtav4`.

Danil Gusak, Nikita Sukhorukov, and Evgeny Frolov. Recommendation is a dish better served warm. In *Proceedings of the Nineteenth ACM Conference on Recommender Systems*, RecSys '25, pp. 1267–1272, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400713644. doi: 10.1145/3705328.3759331. URL `https://doi.org/10.1145/3705328.3759331`.

Moritz Hardt, Eric Mazumdar, Celestine Mendler-Dünner, and Tijana Zrnic. Algorithmic collective action in machine learning. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23, 2023.

F. Maxwell Harper and Joseph A. Konstan. The movielens datasets: History and context. *ACM Trans. Interact. Intell. Syst.*, 5(4), December 2015. ISSN 2160-6455. doi: 10.1145/2827872. URL https://doi.org/10.1145/2827872.

Aditya Karan, Prabhat Kalle, Nicholas Vincent, and Hari Sundaram. Sync or sink: Bounds on algorithmic collective action with noise and multiple groups, 2025a. URL https://arxiv.org/abs/2510.18933.

Aditya Karan, Nicholas Vincent, Karrie Karahalios, and Hari Sundaram. Algorithmic collective action with two collectives. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '25, pp. 1468–1483, New York, NY, USA, 2025b. Association for Computing Machinery. ISBN 9798400714825. doi: 10.1145/3715275.3732098. URL https://doi.org/10.1145/3715275.3732098.

Nadia Karizat, Dan Delmonaco, Motahhare Eslami, and Nazanin Andalibi. Algorithmic folk theories and identity: How tiktok users co-produce knowledge of identity and engage in algorithmic resistance. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2), October 2021. doi: 10.1145/3476046. URL https://doi.org/10.1145/3476046.

Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009. doi: 10.1109/MC.2009.263.

Bo Li, Yining Wang, Aarti Singh, and Yevgeniy Vorobeychik. Data poisoning attacks on factorization-based collaborative filtering. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, pp. 1893–1901, Red Hook, NY, USA, 2016. Curran Associates Inc. ISBN 9781510838819.

M. Olson. *The Logic of Collective Action: Public Goods and the Theory of Groups, With a New Preface and Appendix*. American studies collection. Harvard University Press, 1965. ISBN 9780674537514. URL https://books.google.ca/books?id=jzTeOLtf7_wC.

E. Ostrom. *Governing the Commons: The Evolution of Institutions for Collective Action*. Political Economy of Institutions and Decisions. Cambridge University Press, 1990. ISBN 9780521405997. URL https://books.google.ca/books?id=4xg6oUobMz4C.

Garima Pruthi, Frederick Liu, Satyen Kale, and Mukund Sundararajan. Estimating training data influence by tracing gradient descent. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 19920–19930. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/e6385d39ec9394f2f3a354d9d2b88eec-Paper.pdf.

N. Seaver. *Computing Taste: Algorithms and the Makers of Music Recommendation*. University of Chicago Press, 2022. ISBN 9780226822976. URL https://books.google.ca/books?id=dVuVEAAAQBAJ.

Sulthana Shams and Douglas Leith. Evaluating impact of user-cluster targeted attacks in matrix factorisation recommenders. *ACM Trans. Recomm. Syst.*, 3(2), November 2024. doi: 10.1145/3674157. URL https://doi.org/10.1145/3674157.

G. W. Stewart and Ji-guang. Sun. *Matrix perturbation theory*. Computer science and scientific computing. Academic Press, Boston, 1990. ISBN 0126702306.

Agnideven Palanisamy Sundar, Feng Li, Xukai Zou, Tianchong Gao, and Evan D. Russomanno. Understanding shilling attacks and their detection traits: A comprehensive survey. *IEEE Access*, 8:171703–171715, 2020. doi: 10.1109/ACCESS.2020.3022962.

Nicholas Vincent, Hanlin Li, Nicole Tilly, Stevie Chancellor, and Brent Hecht. Data leverage: A framework for empowering the public in its relationship with technology companies. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, pp. 215–227, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450383097. doi: 10.1145/3442188.3445885. URL https://doi.org/10.1145/3442188.3445885.

Qing Xiao, Yuhang Zheng, Xianzhe Fan, Bingbing Zhang, and Zhicong Lu. Let's influence algorithms together: How millions of fans build collective understanding of algorithms and organize coordinated algorithmic actions. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400713941. doi: 10.1145/3706598.3713279. URL https://doi.org/10.1145/3706598.3713279.

Yang Zhang, Zhiyu Hu, Yimeng Bai, Jiancan Wu, Qifan Wang, and Fuli Feng. Recommendation unlearning via influence function. *ACM Trans. Recomm. Syst.*, 3(2), December 2024. doi: 10.1145/3701763. URL https://doi.org/10.1145/3701763.

# A Limitations & Future Work

## A.1 Theoretical Limitations

We can foresee several limitations of the theoretical and empirical results presented in the main body of our work. First, we note that our theoretical analysis focuses primarily on analyzing the collective action effects of a single item-side update (see Theorem 1) where we consider a setting that holds all other representations fixed and assumes members share a common rating for the target item. While the downstream recommendation metrics used in our empirical results suggest that our theoretical definition for success meaningfully highlights the patterns which persist through the system after refitting (e.g. imperfect cohesion and coordination can still yield meaningful influence for a small collective), we understand that this limits the scope of our theoretical findings.

## A.2 Empirical Limitations

While we do validate our results on real world data, as explained in Section E.2 below, there are concrete challenges of controlling for a specific level of cohesion in datasets based on real ratings. While this motivates our need for synthetic experiments, it also serves as a limitation of our work because our key findings are based mostly on synthetic data (which is then corroborated through real world experiments).

Additionally, since we are not running our experiments in any real, online setting, it is hard to fully simulate the behaviors of a collective. While we attempt to synthesize realistic heuristic strategies, the ones tested are not exhaustive.

# B Notation

Table 1: Summary of notation. Primes denote quantities *after* the collective's strategic action, and $\top$ denotes transpose.

| Symbol | Description | Ref. |
|---|---|---|
| *Users, items, and ratings* | | |
| $\mathcal{U}$, $\mathcal{I}$ | Sets of users and items, with $n = \lvert\mathcal{U}\rvert$ users and $m = \lvert\mathcal{I}\rvert$ items. The user set is fixed throughout; no new users are added. | Section 3.1 |
| $R$ | Partially observed ratings matrix, $R \in \mathbb{R}^{n\times m}$ (rows are users, columns are items), with $R_{ij} \in [1, 5]$. | Section 3.1 |
| $\Omega$ | Set of observed entries $(i, j)$ of $R$. $\Omega_{i,\mathcal{I}}$ is the set of items rated by user $i$; $\Omega_{\mathcal{U},j}$ is the set of users who rated item $j$; $\Omega_{\mathcal{U}\setminus C,j}$ is the subset of those users outside the collective $C$ who rated item $j$. | Section 3.1 |
| *Recommender model (ALS)* | | |
| $d$ | Dimension of the latent space. | Section 3.1 |
| $\lambda$ | Regularization strength ($\lambda > 0$): a ridge penalty on the squared norms of the user and item vectors. Larger $\lambda$ shrinks representations toward zero, and $\lambda > 0$ ensures the matrices inverted in the ALS updates are invertible. | Section 3.1 |
| $I$ | The $d \times d$ identity matrix. | Section 3.1 |
| $U$, $V$ | User and item representation matrices, $U \in \mathbb{R}^{n\times d}$ and $V \in \mathbb{R}^{m\times d}$, with rows $u_i, v_j \in \mathbb{R}^d$. | Section 3.1 |
| $f_\Theta$, $g_\Theta$, $s_\Theta$ | User encoder, item encoder, and scoring function. Under ALS, $f_\Theta$ and $g_\Theta$ are the closed-form updates of Section 3.1, and $s_\Theta(u, v) = u^\top v$. | Section 3.1 |
| $\Theta$ | The recommender model: its fixed specification (e.g., $\lambda$) together with the learned representations $(U, V)$. | Section 3.1 |
| $\hat{r}$ | Predicted ratings, $\hat{r} = u_i^\top v_j$. | Section 3.1 |
| *Collective and strategic action* | | |

*Continued on next page*

*Table 1 (continued)*

| Symbol | Description | Ref. |
|---|---|---|
| $C,\ \alpha$ | The collective $C \subseteq \mathcal{U}$ (existing users) and its share of the user base, $\alpha = \lvert C\rvert/n \in (0,1]$, so that $\lvert C\rvert = \alpha n$. | Section 4.1 |
| $j^\star$ | Target item that the collective seeks to promote. | Section 4.2 |
| $u_h$ | Strategic direction, $u_h = f_\Theta(h) \in \mathbb{R}^d$: the user representation induced by rating according to $h$. | |
| $\bar{u}_C,\ \Sigma_C$ | Mean and covariance of members' prior representations, $\bar{u}_C = \frac{1}{\lvert C\rvert}\sum_{i\in C} u_i$ and $\Sigma_C \in \mathbb{R}^{d\times d}$. Small $\Sigma_C$ means the collective is cohesive before acting. | Equation (1) |
| $\epsilon_i$ | Deviation of member $i$ from the collective mean, $\epsilon_i = u_i - \bar{u}_C$, so that $\sum_{i\in C}\epsilon_i = 0$. | Definition 1 |
| $\beta$ | Coordination, $\beta \in [0,1]$: how far the collective mean moves from $\bar{u}_C$ toward $u_h$ ($\beta = 0$: no movement; $\beta = 1$: full movement). | Section 4.1, Definition 1 |
| $\gamma$ | Heterogeneity retention, $\gamma \in [0,1]$: the fraction of members' initial deviations $\epsilon_i$ that persists after acting ($\gamma = 0$: members are identically dispersed after acting). | Section 4.1, Definition 1 |
| $u_i,\ u_i'$ | Representation of member $i \in C$ before and after strategic action, with $u_i' = (1-\beta)\bar{u}_C + \beta u_h + \gamma\epsilon_i$. | Definition 1 |
| $\bar{u}_C',\ \Sigma_C'$ | Post-action mean and covariance of the collective: $\bar{u}_C' = (1-\beta)\bar{u}_C + \beta u_h$ and $\Sigma_C' = \gamma^2\Sigma_C$. | Equation (3) |
| *Success and its components* | | |
| $\bar{u}$ | Mean representation of all users before strategic action (including members of $C$), $\bar{u} = \frac{1}{\lvert\mathcal{U}\rvert}\sum_{i\in\mathcal{U}} u_i$. | Definition 2 |
| $E_{j^\star}(U,V)$ | Predicted relevance of $j^\star$ averaged over users, $\frac{1}{\lvert\mathcal{U}\rvert}\sum_{i\in\mathcal{U}} u_i^\top v_{j^\star} = \bar{u}^\top v_{j^\star}$. | Definition 2 |
| $v_{j^\star},\ v_{j^\star}'$ | Representation of $j^\star$ before and after strategic action. In the theory, $v_{j^\star}'$ is the result of one item-side ALS update given the post-action user representations. | Section 3.1 and eq. (5). |
| $S_{(\alpha,\beta,\gamma,h)}$ | Collective success, $E_{j^\star}(U,V') - E_{j^\star}(U,V) = \bar{u}^\top(v_{j^\star}' - v_{j^\star})$; | Definition 3 |
| $A,\ b$ | Normal-equation terms for the target item from non-collective raters: $A = \sum_{i\in\Omega_{\mathcal{U}\setminus C,j^\star}} u_i u_i^\top + \lambda I \in \mathbb{R}^{d\times d}$ (positive definite) and $b = \sum_{i\in\Omega_{\mathcal{U}\setminus C,j^\star}} R_{ij^\star} u_i$, so that $v_{j^\star} = A^{-1}b$. Members of $C$ are assumed not to have rated $j^\star$ before acting. | Section 4.3 |
| *Evaluation and experiments* | | |
| $T@K,\ \Delta T@K$ | Fraction of users with $j^\star$ among their top-$K$ items, and its change after strategic action, $T@K(j^\star;\Theta') - T@K(j^\star;\Theta)$. | Section 5.1 |
| $D_C,\ D_{\mathcal{U}},\ D_C^{\mathrm{rel}}$ | Dispersion of the collective, $D_C = \mathrm{tr}(\Sigma_C)$, and of all users, $D_{\mathcal{U}} = \mathrm{tr}(\Sigma_{\mathcal{U}})$ (with $\Sigma_{\mathcal{U}}$ the covariance of all user representations), and relative dispersion $D_C^{\mathrm{rel}} = D_C/D_{\mathcal{U}}$, used to set initial cohesion. | Section E.2 |

## C Omitted Lemmas

**Lemma 1 (Sherman–Morrison Identity).** *Let $A \in \mathbb{R}^{d\times d}$ be an invertible matrix, and let $u, v \in \mathbb{R}^d$. If*

$$1 + v^\top A^{-1} u \neq 0,$$

*then the rank-one perturbation $A + uv^\top$ is invertible and satisfies*

$$(A + uv^\top)^{-1} = A^{-1} - \frac{A^{-1}uv^\top A^{-1}}{1 + v^\top A^{-1}u}.$$

# D Deferred Proofs

## D.1 Collective Action Preliminaries

In this section, we begin by deriving all the formulations for our novel collective action model. The key contribution of our work centers around modeling strategic collective action by separating two properties that are collapsed in conventional unified strategic-user models: *coordination*, the extent to which members move toward a shared strategic direction, and *cohesion*, the extent to which members retain heterogeneous prior representations.

### D.1.1 Prior Collective State

As mentioned in the main body of our work, we begin by reminding the reader of our collective's state prior to their strategic action. Let $C \subseteq \mathcal{U}$ denote a collective. We can define the state of a collective prior to strategic action as:

**Definition 4 (Prior Collective State).** *The prior state of collective $C$ is characterized by its mean representation and covariance. Respectively,*

$$\bar{u}_C = \frac{1}{\alpha n} \sum_{i \in C} u_i, \quad \Sigma_C = \frac{1}{\alpha n} \sum_{i \in C} (u_i - \bar{u}_C)(u_i - \bar{u}_C)^\top.$$

Following the strategic action, defined in Definition 1, we can formally derive the post-action mean and covariance.

**Lemma 2.** *[Post-action mean] Under the strategic action model, the mean representation of the collective after coordination is $\bar{u}'_C = (1-\beta)\bar{u}_C + \beta u_h$.*

*Proof.* By definition, the average collective embedding after strategic action can be represented as $\bar{u}'_C = \frac{1}{\alpha n} \sum_{i \in C} u'_i$. Following the definition of $u'_i$, we can rewrite $\bar{u}'_C$ as

$$\begin{aligned} \bar{u}'_C &= \frac{1}{\alpha n} \sum_{i \in C} \big[(1-\beta)\bar{u}_C + \beta u_h + \gamma \epsilon_i\big] \\ &= \frac{1}{\alpha n} \Big[(1-\beta) \sum_{i \in C} \bar{u}_C + \beta \sum_{i \in C} u_h + \gamma \sum_{i \in C} \epsilon_i\Big]. \end{aligned}$$

By our assumption that $\sum_{i \in C} \epsilon_i = 0$, we can simplify:

$$\begin{aligned} \bar{u}'_C &= \frac{1}{\alpha n} \Big[(1-\beta) \sum_{i \in C} \bar{u}_C + \beta \sum_{i \in C} u_h\Big] \\ &= (1-\beta)\bar{u}_C + \beta u_h. \end{aligned}$$

□

**Lemma 3.** *[Post-action covariance] Under the strategic action model, the covariance of the collective after strategic action satisfies $\Sigma'_C = \gamma^2 \Sigma_C$.*

*Proof.* By Definition 1

$$u'_i = (1-\beta)\bar{u}_C + \beta u_h + \gamma \epsilon_i,$$

where $\epsilon_i = u_i - \bar{u}_C$.

From Lemma 2

$$\bar{u}'_C = (1-\beta)\bar{u}_C + \beta u_h.$$

Therefore,

$$\begin{aligned} u'_i - \bar{u}'_C &= \Big[(1-\beta)\bar{u}_C + \beta u_h + \gamma \epsilon_i\Big] - \Big[(1-\beta)\bar{u}_C + \beta u_h\Big] \\ &= \gamma \epsilon_i \\ &= \gamma (u_i - \bar{u}_C). \end{aligned}$$

Substituting into the covariance definition,

$$\begin{aligned}\Sigma'_C &= \frac{1}{\alpha n}\sum_{i\in C}\left(u'_i - \bar{u}'_C\right)\left(u'_i - \bar{u}'_C\right)^\top \\ &= \frac{1}{\alpha n}\sum_{i\in C}\gamma^2\left(u_i - \bar{u}_C\right)\left(u_i - \bar{u}_C\right)^\top \\ &= \gamma^2\Sigma_C.\end{aligned}$$

□

### D.1.2 Formalizing Success in Collective Action Setting

Now we will derive the success conditions for the general collective action setting. First, recall that we formalize the collective's goal as a target item shift induced by collective action. More formally:

**Definition 2.** *[Average Predicted Relevance] Given user representations $U = \{u_i\}_{i\in\mathcal{U}}$ and a target-item representation $v_{j^\star}$, we define the average predicted relevance of item $j^\star$ as*

$$E_{j^\star}(U,V) = \frac{1}{|\mathcal{U}|}\sum_{i\in\mathcal{U}} s_\Theta(u_i, v_{j^\star}) = \frac{1}{|\mathcal{U}|}\sum_{i\in\mathcal{U}} u_i^\top v_{j^\star} = \bar{u}^\top v_{j^\star},$$

*where $s_\Theta$ is the inner-product scoring function of a recommender and $\bar{u} = \frac{1}{|\mathcal{U}|}\sum_{i\in\mathcal{U}} u_i$ denotes the mean user representation.*

Recall that success is defined (as in the main body of our work) as:

**Definition 3.** *[Collective Success: Target Item Shift] Let $(U,V)$ denote the user and item representations before the collective's strategic action $h$, and let $V'$ denote the item representations after a collective of users engages in strategic action (Definition 1). We say that the collective is* successful *if $E_{j^\star}(U,V') > E_{j^\star}(U,V)$. We measure the magnitude of success as $S_{(\alpha,\beta,\gamma,h)} := E_{j^\star}(U,V') - E_{j^\star}(U,V)$, where $\alpha$, $\beta$, $\gamma$, and $h$ parameterize the characteristics of the collective.*

A collective will succeed if and only if $\bar{u}(v'_{j^\star} - v_{j^\star})$ is positive:

$$S_{(\alpha,\beta,\gamma,h)} > 0 \iff \bar{u}(v'_{j^\star} - v_{j^\star}) > 0.$$

Now, we must characterize each element of this success and then combine them to derive the formal success conditions.

Recall that the ALS update for the target item $j^\star$ is

$$v'_{j^\star} = \left[\sum_{i\in\Omega_{\mathcal{U},j^\star}} u_i u_i^\top + \lambda I\right]^{-1}\left[\sum_{i\in\Omega_{\mathcal{U},j^\star}} R_{ij^\star}u_i\right].$$

Following the collective's strategic action, every collective member assigns the target item the common rating $r_{Cj^\star}$. Recall that in Equation (4) we define

$$A = \sum_{i\in\Omega_{\mathcal{U}\setminus C,j^\star}} u_i u_i^\top + \lambda I$$

and

$$b = \sum_{i\in\Omega_{\mathcal{U}\setminus C,j^\star}} r_{ij^\star}u_i.$$

**Theorem 3.** *[Generalized collective target item update] Suppose that every collective member in $C$ follows the strategic action model and assigns the target item $j^\star$ the same rating $r_{Cj^\star}$. Then the post-strategic action representation of $j^\star$ is*

$$v'_{j^\star} = \left[A + \alpha n\left(\bar{u}'_C\bar{u}'^\top_C + \gamma^2\Sigma_C\right)\right]^{-1}\left[b + \alpha n r_{Cj^\star}\bar{u}'_C\right].$$

*Proof.* We can separate the contributions of the collective $C$ and the remaining users $\mathcal{U} \setminus C$:

$$v'_{j^\star} = \left[A + \sum_{i \in C} u'_i u'^\top_i\right]^{-1} \left[b + \sum_{i \in C} r_{Cj^\star} u'_i\right].$$

We first characterize the collective's contribution to the normal-equations matrix. Recall that

$$u'_i = (1-\beta)\bar{u}_C + \beta u_h + \gamma \epsilon_i,$$

where

$$\sum_{i \in C} \epsilon_i = 0.$$

By Lemma 2,

$$\bar{u}'_C = (1-\beta)\bar{u}_C + \beta u_h.$$

Hence

$$u'_i = \bar{u}'_C + \gamma \epsilon_i.$$

It follows that

$$\begin{aligned}
\sum_{i \in C} u'_i u'^\top_i &= \sum_{i \in C} (\bar{u}'_C + \gamma\epsilon_i)(\bar{u}'_C + \gamma\epsilon_i)^\top \\
&= \alpha n \bar{u}'_C \bar{u}'^\top_C + \gamma \bar{u}'_C \left(\sum_{i \in C} \epsilon_i\right)^\top + \gamma \left(\sum_{i \in C} \epsilon_i\right) \bar{u}'^\top_C + \gamma^2 \sum_{i \in C} \epsilon_i \epsilon_i^\top \\
&= \alpha n \bar{u}'_C \bar{u}'^\top_C + \gamma^2 \sum_{i \in C} \epsilon_i \epsilon_i^\top.
\end{aligned}$$

By definition,

$$\Sigma_C = \frac{1}{\alpha n} \sum_{i \in C} \epsilon_i \epsilon_i^\top.$$

Therefore,

$$\sum_{i \in C} u'_i u'^\top_i = \alpha n \left(\bar{u}'_C \bar{u}'^\top_C + \gamma^2 \Sigma_C\right).$$

Next, since every collective member assigns the target item the same rating $r_{Cj^\star}$,

$$\begin{aligned}
\sum_{i \in C} r_{Cj^\star} u'_i &= r_{Cj^\star} \sum_{i \in C} u'_i \\
&= \alpha n r_{Cj^\star} \bar{u}'_C.
\end{aligned}$$

Substituting these expressions into the ALS update gives

$$v'_{j^\star} = \left[A + \alpha n \left(\bar{u}'_C \bar{u}'^\top_C + \gamma^2 \Sigma_C\right)\right]^{-1} \left[b + \alpha n r_{Cj^\star} \bar{u}'_C\right],$$

which proves the result. □

The collective's strategic action can be seen as a perturbation to the initial matrix $A$, which can be decomposed into the post-action collective mean $\bar{u}'_C$ and its residual heterogeneity $\gamma^2 \Sigma_C$. We can define this perturbation as

$$A' := A + \alpha n \left(\bar{u}'_C \bar{u}'^\top_C + \gamma^2 \Sigma_C\right),$$

Using each of these elements, we can now formally characterize the success with $\bar{u}(v'_{j^\star} - v_{j^\star}) > 0$.

**Theorem 1.** *[Collective-Action Success] For a collective characterized by $\alpha$, $\beta$, and $\gamma$ that follows strategy $h$ inducing strategic direction $u_h$, the population-level success is*

$$S_{(\alpha,\beta,\gamma,h)} = \alpha n \, \bar{u}^\top (A')^{-1} \left[\bar{u}'_C \left(r_{Cj^\star} - \bar{u}'^\top_C v_{j^\star}\right) - \gamma^2 \Sigma_C v_{j^\star}\right].$$

*Proof.* Recall from Theorem 3 that the post-strategic action representation of the target item $j^\star$ is

$$v'_{j^\star} = \left[A + \alpha n \left(\bar{u}'_C \bar{u}'^\top_C + \gamma^2 \Sigma_C\right)\right]^{-1} \left[b + \alpha n r_{Cj^\star} \bar{u}'_C\right].$$

We now characterize the change in the target-item representation as

$$v'_{j^\star} - v_{j^\star} = \left(\left[A + \alpha n \left(\bar{u}'_C \bar{u}'^\top_C + \gamma^2 \Sigma_C\right)\right]^{-1} \left[b + \alpha n r_{Cj^\star} \bar{u}'_C\right]\right) - v_{j^\star}$$

We can rewrite $v_{j^\star}$ using the identity

$$v_{j^\star} = \left[A + \alpha n \left(\bar{u}'_C \bar{u}'^\top_C + \gamma^2 \Sigma_C\right)\right]^{-1} \left[A + \alpha n \left(\bar{u}'_C \bar{u}'^\top_C + \gamma^2 \Sigma_C\right)\right] v_{j^\star}$$

since $\left[A + \alpha n \left(\bar{u}'_C \bar{u}'^\top_C + \gamma^2 \Sigma_C\right)\right]^{-1} \left[A + \alpha n \left(\bar{u}'_C \bar{u}'^\top_C + \gamma^2 \Sigma_C\right)\right] = I$

Therefore,

$$\begin{aligned} v'_{j^\star} - v_{j^\star} = &\left[A + \alpha n \left(\bar{u}'_C \bar{u}'^\top_C + \gamma^2 \Sigma_C\right)\right]^{-1} \left[b + \alpha n r_{Cj^\star} \bar{u}'_C\right] \\ &- \left[A + \alpha n \left(\bar{u}'_C \bar{u}'^\top_C + \gamma^2 \Sigma_C\right)\right]^{-1} \left[A + \alpha n \left(\bar{u}'_C \bar{u}'^\top_C + \gamma^2 \Sigma_C\right)\right] v_{j^\star}. \end{aligned}$$

Factoring out the common inverse,

$$\begin{aligned} v'_{j^\star} - v_{j^\star} = &\left[A + \alpha n \left(\bar{u}'_C \bar{u}'^\top_C + \gamma^2 \Sigma_C\right)\right]^{-1} \\ &\cdot \left[b + \alpha n r_{Cj^\star} \bar{u}'_C - A v_{j^\star} - \alpha n \left(\bar{u}'_C \bar{u}'^\top_C + \gamma^2 \Sigma_C\right) v_{j^\star}\right]. \end{aligned}$$

Recall that before the collective acts, the target-item representation satisfies

$$v_{j^\star} = A^{-1} b,$$

and therefore

$$b = A v_{j^\star}.$$

These terms cancel, yielding

$$\begin{aligned} v'_{j^\star} - v_{j^\star} = &\left[A + \alpha n \left(\bar{u}'_C \bar{u}'^\top_C + \gamma^2 \Sigma_C\right)\right]^{-1} \\ &\cdot \left[\alpha n r_{Cj^\star} \bar{u}'_C - \alpha n \left(\bar{u}'_C \bar{u}'^\top_C + \gamma^2 \Sigma_C\right) v_{j^\star}\right]. \end{aligned}$$

Factoring out $\alpha n$,

$$\begin{aligned} v'_{j^\star} - v_{j^\star} = \alpha n &\left[A + \alpha n \left(\bar{u}'_C \bar{u}'^\top_C + \gamma^2 \Sigma_C\right)\right]^{-1} \\ &\cdot \left[r_{Cj^\star} \bar{u}'_C - \left(\bar{u}'_C \bar{u}'^\top_C + \gamma^2 \Sigma_C\right) v_{j^\star}\right]. \end{aligned}$$

Expanding the matrix-vector product,

$$\left(\bar{u}'_C \bar{u}'^\top_C + \gamma^2 \Sigma_C\right) v_{j^\star} = \bar{u}'_C \left(\bar{u}'^\top_C v_{j^\star}\right) + \gamma^2 \Sigma_C v_{j^\star}.$$

Therefore,

$$\begin{aligned} v'_{j^\star} - v_{j^\star} = \alpha n &\left[A + \alpha n \left(\bar{u}'_C \bar{u}'^\top_C + \gamma^2 \Sigma_C\right)\right]^{-1} \\ &\cdot \left[\bar{u}'_C \left(r_{Cj^\star} - \bar{u}'^\top_C v_{j^\star}\right) - \gamma^2 \Sigma_C v_{j^\star}\right]. \end{aligned}$$

By definition, collective success is

$$S_{(\alpha,\beta,\gamma,h)} = \bar{u}^\top \left(v'_{j^\star} - v_{j^\star}\right).$$

Hence,

$$S_{(\alpha,\beta,\gamma,h)} = \alpha n\, \bar{u}^\top \left[A + \alpha n \left(\bar{u}'_C \bar{u}'^\top_C + \gamma^2 \Sigma_C\right)\right]^{-1} \left[\bar{u}'_C \left(r_{Cj^\star} - \bar{u}'^\top_C v_{j^\star}\right) - \gamma^2 \Sigma_C v_{j^\star}\right].$$

Since $\alpha n > 0$, it follows that

$$S_{(\alpha,\beta,\gamma,h)} > 0$$

if and only if

$$\bar{u}^\top \left[A + \alpha n \left(\bar{u}'_C \bar{u}'^\top_C + \gamma^2 \Sigma_C\right)\right]^{-1} \left[\bar{u}'_C \left(r_{Cj^\star} - \bar{u}'^\top_C v_{j^\star}\right) - \gamma^2 \Sigma_C v_{j^\star}\right] > 0.$$

This establishes the generalized success condition. □

### D.1.3 Unified Clone Setting As a Special Case of Generalized Collective Action Success

We now show that the unified clone setting is a special case of our general collective action model in which users have no prior history and are able to engage in perfectly cohesive and coordinated strategic action. In this setting, all $\alpha n$ collective members of $C$ adopt the same strategic representation $(u_h)$, corresponding to $\beta = 1, \gamma = 0$.

**Proposition 3.** *[Unified collective target item update] In accordance with strategic direction $u_h$, each collective member assigns rating $r_{Cj^\star}$ to the target item $j^\star$. The representation of the target item after strategic action becomes*

$$v'_{j^\star} = \left(A + \alpha n u_h u_h^\top\right)^{-1} \left(b + \alpha n r_{Cj^\star} u_h\right).$$

Under complete cohesion and coordination, every member of the collective is represented by the same strategic direction $u_h$. Since the collective contains $\alpha n$ users and each assigns rating $r_{Cj^\star}$ to $j^\star$, their contribution to the normal matrix is

$$\sum_{i \in C} u_h u_h^\top = \alpha n \, u_h u_h^\top,$$

while their contribution to the response vector is

$$\sum_{i \in C} r_{Cj^\star} u_h = \alpha n \, r_{Cj^\star} u_h.$$

Therefore, after the strategic action, the ALS update for the target item is

$$v'_{j^\star} = \left(A + \alpha n u_h u_h^\top\right)^{-1} \left(b + \alpha n r_{Cj^\star} u_h\right),$$

which proves the result.

Using this formal definition for $v'_{j^*}$ enables us to characterize the success of a fully unified clone collective.

**Corollary 1.** *[Unified-Clone Success] For $\beta = 1$ and $\gamma = 0$,*

$$S_{(\alpha,1,0,h)} = \frac{\alpha n (\bar{u}^\top A^{-1} u_h)(r_{Cj^\star} - u_h^\top v_{j^\star})}{1 + \alpha n u_h^\top A^{-1} u_h}.$$

*Proof.* After the strategic action, each of the $\alpha n$ collective members contributes the same outer product $u_h u_h^\top$ and the same rating-weighted vector $r_{Cj^\star} u_h$ to the ALS normal equations. Hence,

$$A' = A + \alpha n u_h u_h^\top \text{ and } b' = b + \alpha n r_{Cj^\star} u_h.$$

The post-strategic action item representation is therefore

$$v'_{j^\star} = \left(A + \alpha n u_h u_h^\top\right)^{-1} \left(b + \alpha n r_{Cj^\star} u_h\right).$$

Applying the Sherman–Morrison identity,

$$\left(A + \alpha n u_h u_h^\top\right)^{-1} = A^{-1} - \frac{\alpha n A^{-1} u_h u_h^\top A^{-1}}{1 + \alpha n u_h^\top A^{-1} u_h}.$$

Substituting this expression into the post-strategic action update and using $v_{j^\star} = A^{-1} b$ gives

$$\begin{aligned} v'_{j^\star} = v_{j^\star} &+ \alpha n r_{Cj^\star} A^{-1} u_h \\ &- \frac{\alpha n A^{-1} u_h u_h^\top v_{j^\star} + \alpha^2 n^2 r_{Cj^\star} A^{-1} u_h u_h^\top A^{-1} u_h}{1 + \alpha n u_h^\top A^{-1} u_h}. \end{aligned}$$

Collecting terms yields

$$v'_{j^\star} = v_{j^\star} + \frac{\alpha n A^{-1} u_h}{1 + \alpha n u_h^\top A^{-1} u_h} \left(r_{Cj^\star} - u_h^\top v_{j^\star}\right),$$

and hence

$$v'_{j^\star} - v_{j^\star} = \frac{\alpha n A^{-1} u_h}{1 + \alpha n u_h^\top A^{-1} u_h} \left(r_{Cj^\star} - u_h^\top v_{j^\star}\right).$$

Therefore,

$$S_{(\alpha,1,0,h)} = \bar{u}^\top (v'_{j^*} - v_{j^*}) = \frac{\alpha n (\bar{u}^\top A^{-1} u_h)(r_{Cj^\star} - u_h^\top v_{j^\star})}{1 + \alpha n u_h^\top A^{-1} u_h}.$$

□

### D.2 Properties of Success

Given the two definitions of generalized and unified collective action success, we now wish to understand the dynamics relating to increasing the proportion $\alpha$. In Section D.2.1 we use the success formulation $S_{(\alpha,\beta=1,\gamma=0,h)}$ to characterize its relationship with $\alpha$. Following this, in Section D.2.2, we relax the parameters of the simple setting and characterize the more complex behaviors of a true, user-based collective.

#### D.2.1 Under Unified Setting

First, we show that, under the unified clone setting, the success is monotonically increasing in $\alpha$.

**Proposition 1.** *[Monotonicity of Success Under Unified Clone Setting] Assuming that $\bar{u}^\top A^{-1} u_h (r_{Cj^\star} - u_h^\top v_{j^\star}) > 0$, the collective success $S_{(\alpha,\beta=1,\gamma=0,h)}$ is monotonically increasing in $\alpha \in (0, 1]$.*

*Proof.* Recall that

$$S_{(\alpha,\beta=1,\gamma=0,h)} = \bar{u}^\top (v'_{j^*} - v_{j^*}) = \frac{\alpha n (\bar{u}^\top A^{-1} u_h)(r_{Cj^\star} - u_h^\top v_{j^\star})}{1 + \alpha n u_h^\top A^{-1} u_h} = \frac{\alpha n (\bar{u}^\top A^{-1} \bar{u}'_C)(r_{Cj^\star} - \bar{u}_C'^\top v_{j^\star})}{1 + \alpha n \bar{u}_C^\top A^{-1} \bar{u}_C}$$

because when $\beta = 1$ and $\gamma = 0$, $\bar{u}'_C = u_h$.

Let

$$c := \bar{u}^\top A^{-1} u_h, \quad q := u_h^\top A^{-1} u_h, \quad \Delta := r_{Cj^\star} - u_h^\top v_{j^\star}.$$

For simplicity, we can rewrite

$$S_{(\alpha,1,0,h)} = \frac{\alpha n c \Delta}{1 + \alpha n q}.$$

Differentiating with respect to $\alpha$ yields

$$\frac{dS}{d\alpha} = \frac{nc\Delta}{(1 + \alpha n q)^2}.$$

Since, by assumption, $c\Delta > 0$, and we assume $\alpha n > 0$, it holds that

$$\frac{dS}{d\alpha} > 0.$$

Thus $S_{(\alpha,\beta=1,\gamma=0,h)}$ is monotonically increasing in $\alpha$.

□

Here, $c$ captures the alignment between the average user direction and the collective direction under the geometry induced by $A^{-1}$, $q$ measures the self-alignment or strength of the collective direction, and $\Delta$ measures how much higher the collective's target rating is than the model's current prediction for the target item.

We now further examine the monotonic relationship between $\alpha$ and $S_{(\alpha,\beta=1,\gamma=0,h)}$. We show that, in the unified clone setting, success exhibits diminishing returns with respect to the collective proportion $\alpha$. Intuitively this means that each additional member of the collective contributes less to the overall success as the proportion of the collective grows.

**Proposition 2.** *[Diminishing Returns Under Unified Clone Setting] Assuming that* $\bar{u}^\top A^{-1} u_h (r_{Cj^\star} - u_h^\top v_{j^\star}) > 0$*,* $S_{(\alpha,\beta=1,\gamma=0,h)}$ *is strictly increasing and strictly concave in* $\alpha$*. In other words, collective action exhibits diminishing marginal returns in its success with respect to collective size* $\alpha$*.*

*Proof.* Differentiating $S_{(\alpha,\beta=1,\gamma=0,h)}$ with respect to $\alpha$ yields

$$\frac{dS}{d\alpha} = \frac{nc\Delta}{(1+\alpha nq)^2}.$$

Since $n > 0$, $q = u_h^\top A^{-1} u_h > 0$, and by assumption $c\Delta > 0$, it follows that $\frac{dS}{d\alpha} > 0$, so $S_{(\alpha,\beta=1,\gamma=0,h)}$ is strictly increasing. Differentiating once more,

$$\frac{d^2S}{d\alpha^2} = -\frac{2n^2cq\Delta}{(1+\alpha nq)^3}.$$

The denominator is strictly positive and all factors in the numerator are positive except for the leading minus sign. Therefore,

$$\frac{d^2S}{d\alpha^2} < 0.$$

Thus, $S$ is strictly concave in $\alpha$. Strict concavity implies that the marginal gain from increasing collective size *decreases* as $\alpha$ grows, establishing diminishing returns. □

### D.2.2 Under Generalized Setting

We now characterize the properties of $S_{(\alpha,\beta,\gamma,h)}$ with respect to $\alpha$ in the generalized setting. We first find that, if $\gamma = 0$, we recover the same findings of monotonicity of success and diminishing returns as shown in the unified clone setting above.

**Theorem 2.** *[Monotonicity of Success and Diminishing Returns with* $\gamma = 0$*] Assuming* $(\bar{u}^\top A^{-1} \bar{u}'_C)(r_{Cj^\star} - \bar{u}_C'^\top v_{j^\star}) > 0$*, for any constant value* $\beta$*, if* $\gamma = 0$ *the collective success* $S_{(\alpha,\beta,\gamma=0,h)}$ *is monotonically increasing in* $\alpha \in (0,1]$ *and is strictly concave in* $\alpha$*.*

*Proof.* Holding $\beta$ fixed, recall that the coordinated collective mean is defined by

$$\bar{u}'_C = (1-\beta)\bar{u}_C + \beta u_h.$$

Recall that after strategic action, each collective member can be represented as

$$u'_i = (1-\beta)\bar{u}_C + \beta u_h + \gamma\epsilon_i$$

Since $\gamma = 0$, they all have the same representation, so $u'_i = \bar{u}'_C$, and the post-action heterogeneity of the collective is $\Sigma'_C = \gamma^2 \Sigma_C = 0$.

The target update then, once again, becomes a rank-1 perturbation on the initial matrix $A$ such that the difference in the representation of target item before and after its update is (similarly to Proposition 1 and Proposition 2) becomes:

$$v'_{j^\star} - v_{j^\star} = \frac{\alpha n A^{-1} \bar{u}'_C}{1 + \alpha n \bar{u}_C'^\top A^{-1} \bar{u}'_C} \left(r_{Cj^\star} - \bar{u}_C'^\top v_{j^\star}\right).$$

Thus, the success conditions from Propositions 1 and 2 hold.

□

Next, we show that when $\gamma > 0$, monotonicity of success and diminishing returns are not guaranteed to hold in the general case. In particular, we find that the homogeneity (or cohesiveness) of a collective plays a central role in controlling whether or not success is achieved as $\alpha$ increases. Intuitively, we can consider how adding heterogeneous members who decrease the cohesiveness of a collective can deteriorate its ability to achieve success rather than increase it.

**Corollary 2.** *[Monotonicity of Success in the General Case] For $\gamma > 0$, monotonicity of success is not guaranteed to hold in the general case.*

*Proof.* Consider a setting with $n = 20$ users and $\beta = 0$, $\gamma = 1$, (i.e., fully uncoordinated and heterogeneous) and $r_{Cj^\star} = 5$. Suppose that three users (who have not yet rated $j^\star$) have an embedding $(u_i)$ $\begin{bmatrix} 1 \\ 1/\sqrt{2} \end{bmatrix}$, and three additional users (who also have not yet rated $j^\star$ have an embedding $(u_i)$ of $\begin{bmatrix} 1 \\ -1/\sqrt{2} \end{bmatrix}$. For $t \in \{2, 4, 6\}$, let $C_t$ correspond to collectives containing the first $t$ of these users. The collectives are nested, and each has

$$\bar{u}_{C_t} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \Sigma_{C_t} = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}. \tag{*}$$

We first specify the remaining user representations to show that the baseline ALS quantities in this example are realizable. Set $\lambda = 9/25$. Four (non-collective) users have previously rated $j^\star$, each assigning it rating 5. Let their representations be

$$\begin{bmatrix} 4/5 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1/5 \end{bmatrix}, \begin{bmatrix} 0 \\ \sqrt{3/10} \end{bmatrix}, \begin{bmatrix} 0 \\ -\sqrt{3/10} \end{bmatrix}.$$

Suppose further that the other ten users have representation $\begin{bmatrix} 33/25 \\ 99/50 \end{bmatrix}$ and have not rated $j^\star$. Summing over all 20 users gives $\bar{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

From the four (non-collective) users who have already rated $j^\star$, we get

$$A = \sum_{i \in \Omega_{U \setminus C_t, j^\star}} u_i u_i^\top + \lambda I_2 = \frac{16}{25} I_2 + \frac{9}{25} I_2 = I_2 \text{ and } b = \sum_{i \in \Omega_{U \setminus C_t, j^\star}} r_{ij^\star} u_i = \begin{pmatrix} 4 \\ 1 \end{pmatrix}.$$

Consequently, the pre-action target representation is $v_{j^\star} = A^{-1} b = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$.

From (*) we obtain

$$\bar{u}_{C_t} \bar{u}_{C_t}^\top + \Sigma_{C_t} = \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix}$$

and

$$\bar{u}_{C_t} \left(5 - \bar{u}_{C_t}^\top v_{j^\star}\right) - \Sigma_{C_t} v_{j^\star} = \begin{bmatrix} 1 \\ -\frac{1}{2} \end{bmatrix}.$$

Writing $t = |C_t| = \alpha_t n$, Theorem 1 therefore implies that

$$\begin{aligned} S(\alpha_t, 0, 1, h) &= \alpha n \, \bar{u}^\top \left[A + \alpha n \left(\bar{u}'_C \bar{u}'^\top_C + \Sigma_{C_t}\right)\right]^{-1} \left[\bar{u}'_C \left(5 - \bar{u}'^\top_C v_{j^\star}\right) - \Sigma_C v_{j^\star}\right] \\ &= t \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} \frac{1}{1+t} & 0 \\ 0 & \frac{1}{1+t/2} \end{bmatrix} \begin{bmatrix} 1 \\ -\frac{1}{2} \end{bmatrix} \\ &= \frac{t}{1+t} - \frac{t}{2+t} \\ &= \frac{t}{(t+1)(t+2)}. \end{aligned}$$

Then,

$$S(\alpha_2, 0, 1, h) = \frac{1}{6} \quad \text{and} \quad S(\alpha_4, 0, 1, h) = \frac{2}{15}.$$

Both quantities are positive, but $C_2 \subset C_4$ and $1/6 > 2/15$. Thus increasing the collective from two to four members decreases its success.

For comparison (and to validate Theorem 2, setting $\gamma = 0$ while keeping the same setup results in

$$S(\alpha_t, 0, 0, h) = \frac{t}{1+t}.$$

Thus, the corresponding perfectly cohesive collective has strictly increasing success. □

Crucially, this shows that *even in the setting where a collective gives their target item the highest possible score*, their heterogeneity (or lack of cohesiveness) hinders their ability to achieve success.

**Corollary 3.** *[Diminishing Returns in the General Case] For $\gamma > 0$, diminishing returns is not guaranteed to hold in the general case.*

*Proof.* Use the population, strategy, and nested collectives $C_2 \subset C_4 \subset C_6$ constructed in the preceding proof. Their sizes are equally spaced, and $\alpha_4 = (\alpha_2 + \alpha_6)/2$. The success formula derived there gives

$$S(\alpha_2, 0, 1, h) = \frac{1}{6}, \qquad S(\alpha_4, 0, 1, h) = \frac{2}{15}, \qquad S(\alpha_6, 0, 1, h) = \frac{3}{28}.$$

All three values are positive. Nevertheless,

$$\frac{S(\alpha_2, 0, 1, h) + S(\alpha_6, 0, 1, h)}{2} - S(\alpha_4, 0, 1, h) = \frac{1}{280} > 0.$$

Hence success at the midpoint $\alpha_4$ is strictly below the average success at $\alpha_2$ and $\alpha_6$, contradicting concavity. All three collective covariances in this comparison are realized by their respective finite sets of users. □

# E Experimental Details

## E.1 Synthetic Data Generation

We construct a synthetic recommendation environment in which we can independently manipulate the characteristics of the strategic collective. Each environment contains $n = 100$ users and $m = 100$ items distributed across $G = 5$ latent genres. Each genre $g$ is associated with basis vector $e_g \in \mathbb{R}^5$, and item representations are sampled as

$$v_j^{\text{true}} \sim \mathcal{N}\left(e_{g(j)}, \sigma_V^2 I\right), \qquad \sigma_V = 0.2.$$

Items are distributed evenly across genres.

To introduce differences in item popularity, genres are assigned popularity weights

$$\pi = (0.35, 0.25, 0.18, 0.13, 0.09).$$

For each user, we sample baseline interactions in two stages: an observation process that determines which items the user encounters, followed by a rating process that determines how the user evaluates those observed items.

Let

$$s_{ij} = u_i^\top v_j$$

denote the latent affinity between user $i$ and item $j$. We first define an observation distribution over items as

$$P(j \mid i) \propto p_j^\tau \exp(\eta s_{ij}),$$

where $p_j$ denotes the popularity of item $j$, $\tau$ controls the strength of popularity bias, and $\eta$ controls the extent to which observations are driven by the user's latent preferences. When $\tau = 0$, popularity plays no role in the observation process and exposure is determined entirely by user preference. As $\tau$ increases, relatively popular items receive increasingly greater probability of being observed. Similarly, larger values of $\eta$ make users increasingly selective with respect to relevance, concentrating more probability mass on items with larger affinity scores.

Table 2: Synthetic environment hyperparameters used in our experiments.

| Category | Parameter | Value |
|---|---|---|
| Environment | Number of users $\lvert\mathcal{U}\rvert$ | 100 |
| | Number of items $\lvert\mathcal{I}\rvert$ | 100 |
| | Number of genres $G$ | 5 |
| | Items per genre | 20 |
| | Latent dimension $d$ | 5 |
| User / Item Generation | Item noise $\sigma_V$ | 0.2 |
| | User mean | 0.0 |
| | User standard deviation | 1.0 |
| Popularity | Item popularity noise $\sigma_{\text{pop}}$ | 0.5 |
| | Genre popularity | $(0.35, 0.25, 0.18, 0.13, 0.09)$ |
| Interaction Sampling | Interactions per user | 20 |
| | Popularity bias $\tau$ | 1.0 |
| | Preference selectivity $\eta$ | 1.0 |
| Initial Ratings | Rating mean | 3.0 |
| | Minimum rating | 1.0 |
| | Maximum rating | 5.0 |
| | Rating scale | 0.75 |
| | Rating noise $\sigma_R$ | 0.35 |
| Target | Target item $j^\star$ | 35 |
| | Target rating $r_{Cj^\star}$ | 5 |

Using this distribution, we sample $K = 20$ observed items for each user and define the corresponding observation mask $\Omega$. Separately, we generate a potential rating for every user–item pair according to

$$\tilde{r}_{ij} = \text{clip}\left(\mu_R + s_{ij} + \epsilon_{ij}, 1, 5\right), \qquad \epsilon_{ij} \sim \mathcal{N}(0, \sigma_R^2),$$

where $\mu_R$ controls the average rating level and $\sigma_R$ introduces rating noise.

Finally, we apply the observation mask to the complete matrix of simulated ratings:

$$R_{ij} = \begin{cases} \tilde{r}_{ij}, & (i,j) \in \Omega, \\ \text{unobserved}, & \text{otherwise.} \end{cases}$$

For a comprehensive list of the hyperparameters used to characterize our synthetic environment, please see Table 2.

### E.2 Collective Construction

For each synthetic environment, we construct collectives with controlled size and initial heterogeneity. Collective size is determined by $\alpha$, such that

$$|C| = \alpha|\mathcal{U}|.$$

To vary initial collective cohesion, we construct collectives at $tr(\Sigma_C) \sim (0, 0.5, 1.0)$ levels of initial dispersion. For each collective, we record its pre-action mean and covariance,

$$\bar{u}_C = \frac{1}{|C|}\sum_{i \in C} u_i, \qquad \Sigma_C = \frac{1}{|C|}\sum_{i \in C}(u_i - \bar{u}_C)(u_i - \bar{u}_C)^\top.$$

To turn this quantity into an experimentally controllable parameter, we work with the corresponding scalar dispersion,

$$D_C = \text{tr}(\Sigma_C),$$

and compare it to the dispersion of the full user population,

$$D_U = \text{tr}(\Sigma_U).$$

Rather than controlling the raw collective dispersion directly, we control its dispersion relative to the population,

$$D_{\text{rel}} = \frac{D_C}{D_U}.$$

This normalization makes cohesion interpretable across experimental settings, since the magnitude of an unnormalized covariance depends on the scale of the latent space. For example, $D_{\text{rel}} = 0$ corresponds to a perfectly cohesive collective whose members have identical representations, $D_{\text{rel}} = 0.25$ denotes a collective whose dispersion is one quarter that of the overall population, and $D_{\text{rel}} = 1$ denotes a collective that is as dispersed as the population as a whole.

To construct a collective with a desired value of $D_{\text{rel}}$, we first compute the empirical dispersion of the general population $D_U$. Given the experimental parameter $D_{\text{rel}}$, the desired collective dispersion is then

$$D_C = D_{\text{rel}} D_U.$$

We next sample a set of $|C| = \alpha n$ users to form the collective and define a target collective mean $\bar{u}_C$. Let the initial deviations of the selected users around this mean be

$$\epsilon_i = u_i^{(0)} - \bar{u}_C.$$

We compute their current dispersion as

$$D_{\text{current}} = \frac{1}{|C|} \sum_{i \in C} \|\epsilon_i\|_2^2.$$

Because the trace of the covariance is equal to the average squared distance from the mean, we can obtain the desired collective dispersion by rescaling these deviations. Specifically, we replace each collective representation by

$$u_i = \bar{u}_C + \sqrt{\frac{D_C}{D_{\text{current}}}}\, \epsilon_i.$$

This transformation preserves the relative geometry of the sampled collective around $\bar{u}_C$ while uniformly expanding or contracting its deviations so that the resulting collective satisfies

$$\text{tr}\left(\Sigma_C\right) = D_C = D_{\text{rel}} D_U.$$

This construction differs from generating entirely new isotropic noise around $\bar{u}_C$. Instead, it preserves the directions of the selected users' pre-existing deviations and changes only their overall scale. As a result, $D_C^{\text{rel}}$ controls the collective's initial cohesion while retaining the sampled collective's underlying representational structure.

We note that one of the key reasons for using a synthetic dataset is that our experiments rely on our ability to exactly control for a specific desired prior heterogeneity (and cohesion). In an attempt to recreate this synthesis on our real world data (where we do not directly construct latent representations that can align with our theoretical assumptions), we considered three levels of prior heterogeneity, controlled by three sampling techniques inspired by (Karan et al., 2025b). First, we clustered our users into 10 different clusters. A "low" heterogeneity paradigm involved sampling collective members from a single cluster. A "medium" level of heterogeneity consisted of sampling from three different clusters. Finally, a "high" level of prior heterogeneity was associated with sampling randomly from the entire user pool. In Table 3, we can see the realized dispersion, represented as $tr(\Sigma_C)$ associated with a proportion $\alpha$ and the associated heterogeneity level. We can see that as the proportion grows, it becomes increasingly difficult to guarantee that a "low"-level heterogeneity is adequately cohesive, thus motivating our need for synthetic experimentation.

This construction allows us to vary the collective's pre-existing composition independently from the coordination and residual heterogeneity imposed during strategic action. However, it's important to note that, as shown in Table 3, since the heterogeneity of the collective is controlled by the underlying data distribution, for a sufficiently large $\alpha$ it is impossible to cleanly control the cohesion parameter.

Table 3: Realized relative collective dispersion in MovieLens across 100 collective samples.

| $\alpha$ | Low ($tr(\Sigma_C) = 0$) | Medium ($tr(\Sigma_C) = 0.5$) | High ($tr(\Sigma_C) = 1$) |
|---|---|---|---|
| 0.01 | $0.54 \pm 0.21$ | $0.73 \pm 0.28$ | $0.93 \pm 0.29$ |
| 0.05 | $0.60 \pm 0.15$ | $0.85 \pm 0.16$ | $0.99 \pm 0.13$ |
| 0.10 | $0.61 \pm 0.14$ | $0.84 \pm 0.13$ | $1.00 \pm 0.11$ |
| 0.30 | $0.72 \pm 0.17$ | $0.86 \pm 0.12$ | $0.99 \pm 0.05$ |

### E.3 From Theoretical Representation Formulation to Strategic Action

Our theoretical model specifies strategic action directly in representation space. In practice, however, users cannot directly choose their learned representations; they can only modify their observable interaction histories. We therefore distinguish between the *theoretical* collective state specified by the theory and the *realized* state induced by feasible strategic behavior.

For each collective member, the theoretical post-action representation is

$$u_i'^{(\text{theory})} = (1-\beta)\bar{u}_C + \beta u_h + \gamma\epsilon_i, \qquad \epsilon_i = u_i - \bar{u}_C,$$

as previously defined in Definition 1.

Given these desired representations, we construct feasible modifications to each member's rating history using the strategic behavior induced by $h$. These modifications produce a post-action ratings matrix $R'$.

Before retraining the recommender, we pass each modified history through the *frozen baseline user encoder*,

$$u_i'^{(\text{real})} = f_\Theta(R_i'),$$

and use these representations to quantify the intervention actually realized in representation space. Only after this measurement do we fine-tune/refit the recommender on the strategic interaction data.

Thus, our experimental pipeline is

$$U_C \longrightarrow U_C'^{(\text{theory})} \longrightarrow R' \longrightarrow U_C'^{(\text{real})} \longrightarrow U_C'^{\text{refit}},$$

separating the requested intervention from both its behavioral realization and the subsequent adaptation of the recommender.

### E.4 Collective Strategy Construction

Each strategy $h$ specifies a target-item rating and a set of additional "filler" interactions intended to induce a strategic direction $u_h$. Unless otherwise stated, collective members assign the maximum rating $r_{Cj^\star} = 5$ to the target item.

**Random Filler.** We sample filler items uniformly at random from the set of eligible items. This provides a strategy with no explicit relationship to either the target item or the broader population.

**Popular-Item Mimic.** We rank candidate filler items according to their popularity in the baseline interaction data and select the most popular eligible items. This strategy moves collective behavior toward broadly prevalent interaction patterns.

**Target Neighborhood.** We cluster baseline item representations and identify the cluster containing the target item $j^\star$. Filler items are selected from this cluster, producing strategic behavior concentrated around the target's local representation-space neighborhood.

**Target-Supporter Mimic.** We cluster users according to their baseline representations and identify the cluster exhibiting the strongest support for $j^\star$. We then select filler items preferred by members of this cluster. This strategy attempts to imitate the behavior of users whose preferences are already favorable to the target.

**Filler Item Budget.** Each collective member is allowed a fixed strategic-action budget of $B$ interactions. Our primary experiments use $B = 10$.

### E.5 Optimal Latent-Space Strategy

In addition to behaviorally realizable heuristics, we construct an optimal benchmark that directly optimizes the strategic direction $u_h$ according to the theoretical Item-Shift Success objective. For fixed $(\alpha, \beta, \gamma)$ and prior collective state $(\bar{u}_C, \Sigma_C)$, we solve

$$u_h^\star = \arg\max_{u_h} S_{(\alpha,\beta,\gamma,u_h)}.$$

**Unbounded Optimally Solved Strategic Direction.** The unbounded oracle places no behavioral constraint on the optimized latent direction. It therefore represents an idealized upper benchmark for the theoretical objective rather than a realizable user strategy.

**Movement-Matched Optimally Solved Strategic Direction.** For comparison with a behavioral heuristic $h$, we additionally constrain the oracle to use the same latent-space movement budget as the direction induced by that heuristic:

$$u_h^* = \arg\max_{u_h} S_{(\alpha,\beta,\gamma,u_h)} \quad \text{s.t.} \quad \|u_h - \bar{u}_C\|_2 = \|u_h^* - \bar{u}_C\|_2 .$$

This benchmark isolates whether differences between a heuristic and the theoretical optimum arise from the direction of strategic movement rather than simply its magnitude.

### E.6 Recommender Models and Training

#### E.6.1 Alternating Least Squares

We use matrix factorization with squared-error reconstruction and $\ell_2$ regularization as the model corresponding directly to our theoretical analysis. Given observed interactions $\Omega$, ALS alternates between the closed-form user and item updates described in Section 3.1.

For each experimental condition, we first fit a baseline model to $R$. Following strategic action, we warm-start from the baseline parameters and refit on the augmented interaction data $R'$. Please note that because the synthetic ALS experiments are designed to provide a controlled empirical validation of the theoretical dynamics, ALS is fit to the complete simulated interaction matrix rather than optimized for held-out recommendation performance. For our real-world experiments on the MovieLens 100K dataset, we use the traditional train/validation/test of 80/10/10 splits. Our model uses a regularization of $\lambda = 10$, 15 rounds of fitting, and an early stopping patience of 3 epochs without validation metric improvement, and hidden dimension of 64. Our model achieves a pre-action Recall@10 of 0.02, RMSE of 0.987 and a post-action Recall@10 of 0.03. We understand that this is below SOTA but since the goal of our method is not to make assessments about the quality of the recommendations, we accept this as a valid base from which to make our evaluations.

### E.7 Evaluation Metrics

We distinguish three levels at which strategic action may affect the target item.

**Item-Shift Success.** Our theoretical metric isolates movement of the target item while evaluating that movement relative to the baseline population:

$$S_{\text{target item shift}} = \bar{u}^\top \left(v'_{j^\star} - v_{j^\star}\right).$$

This corresponds directly to the success quantity analyzed in Theorem 1.

**Post-Refit Success.** After retraining the recommender, both user and item representations may change. We therefore measure the average change in predicted relevance of the target item:

$$S_{\text{refit}} = \frac{1}{|\mathcal{U}|} \sum_{i \in \mathcal{U}} \left[ (u_i'^{\text{refit}})^\top v'_{j^\star} - u_i^\top v_{j^\star} \right].$$

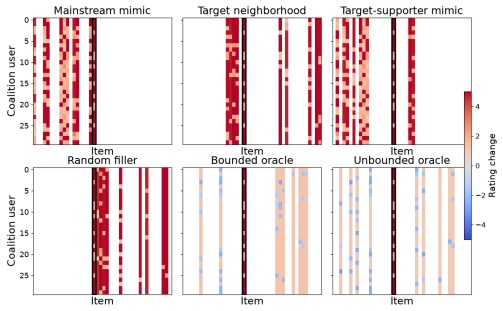



Figure 5: Realized rating modifications $R'$ under the (bounded) optimized and several selected heuristic strategies, shown for $\alpha = 0.3$, $\beta = 1.0$, and $\gamma = 0.0$. Rows correspond to collective members and columns to items. Whereas the realizable heuristics primarily add high ratings to previously unrated strategic items, the optimized strategy additionally reduces ratings assigned to previously rated items. Note: the figure is repeated from main body for clarity.

**Top-$K$ Change.** Finally, we measure whether representational and score changes translate into recommendation exposure. Define

$$T@K(j^\star;\Theta) = \frac{1}{|\mathcal{U}|}\sum_{i\in\mathcal{U}} \mathbb{I}\left[j^\star \in \mathrm{TopK}_i(\Theta)\right].$$

We report

$$\Delta T@K = T@K(j^\star;\Theta') - T@K(j^\star;\Theta).$$

Together, these metrics distinguish immediate target-item movement, post-retraining relevance change, and final ranking exposure.

# F Additional Results

## F.1 Full Ratings Visualization

In this section we visualize the effects of applying our heuristic and optimal strategies to a collective. For each of the heuristic strategies (detailed in Appendix E.4) we first select a set of strategic items and their respective strategic ratings. Following the process described in Appendix E.3, we then solve for the strategic direction vector $u_h$ induced by following this particular heuristic. This involves solving the ALS normal equations for the resulting user update if they had no prior history and explicitly rated only the strategic item set with their associated strategic ratings. Meanwhile, in the optimal case, we use the solver detailed in Appendix E.5 to solve for an optimal strategic direction $u_h$. We then synthesize the strategically modified ratings matrix $R'$ by modifying the entries $R_{ij}$ for $i \in C, j \in$ strategic item set. Below, in Figure 5, we can see the ratings realization of each of these strategies. More importantly, we can see that our heuristic strategies follow a realistic, additive protocol, where users increase ratings for items following their heuristic strategy. Meanwhile, the optimal strategies function by lowering previously rated items.

## F.2 Synthetic Hyperparameter Sweeps

To complement the main synthetic results, we conduct a broader sweep over the collective-action parameters to examine whether the observed effects persist across experimental configurations. In particular, we vary collective size $\alpha$, coordination $\beta$, and initial collective dispersion $\mathrm{tr}(\Sigma_C)$ across each behavioral strategy. Figure 6 summarizes the resulting success surfaces.

## F.3 Applying ALS to MovieLens

Below, we recreate the synthetic experimental setup in the main body of our work by replacing a simulated ground truth ratings set $R$ with that of a real movie recommendation dataset called *Movie-*

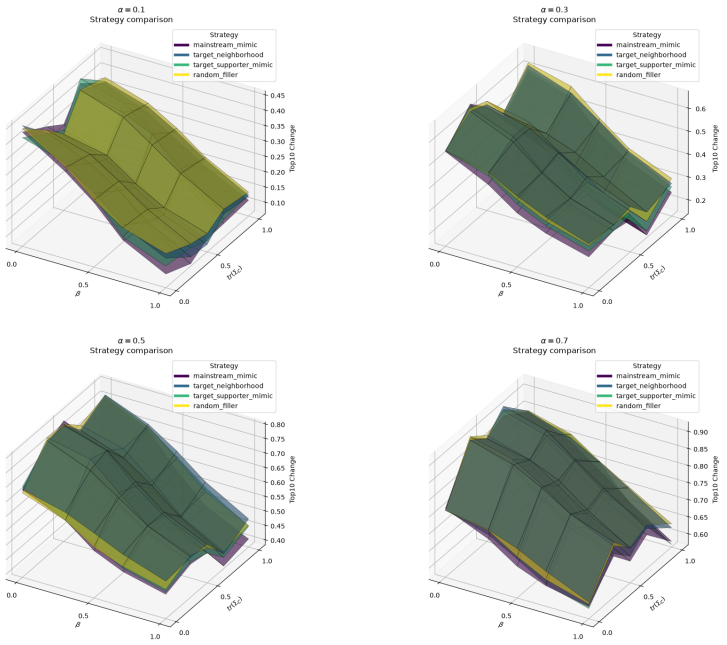



Figure 6: We simulate all the permutations of cohesion/coordination parameters which can be seen in our synthetic ALS experiments. From these plots we can see that irrespective of the heuristic strategy, a higher proportion $\alpha$ of the population joining a collective manifests in increased success.

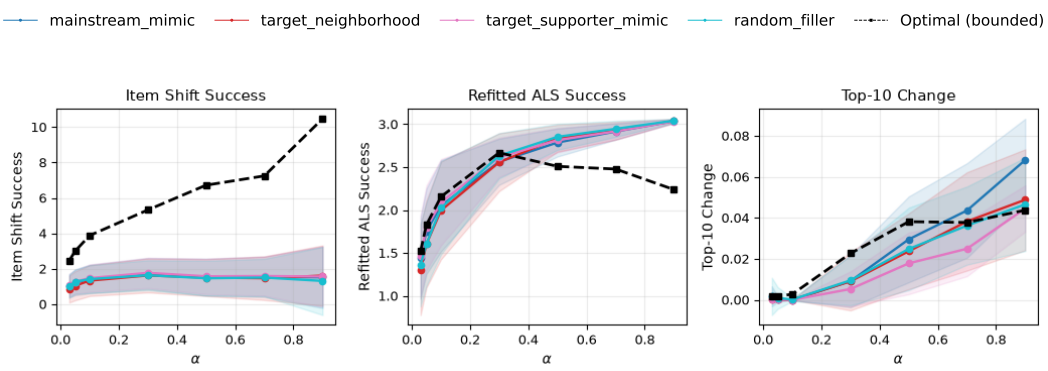
*Lens 100K* (Harper & Konstan, 2015). Before diving into those results, we present the statistics of our synthesized data and the corresponding value for MovieLens.

Table 4: Comparison of the synthetic environment and MovieLens-100K.

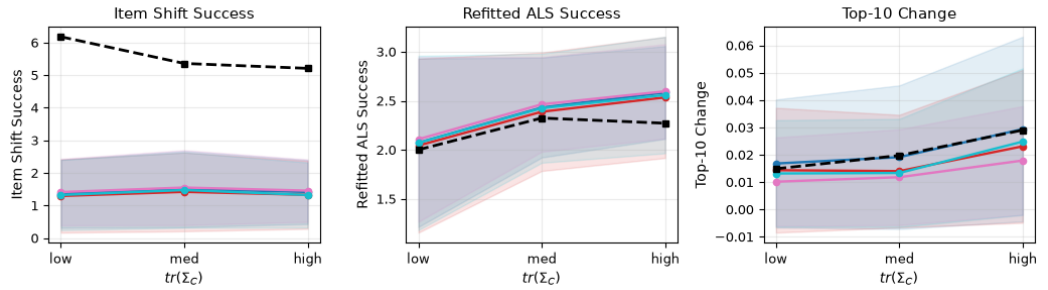
| | **Synthetic** | **MovieLens-100K** |
|---|---|---|
| Users | 100 | 943 |
| Items | 100 | 1682 |
| Interactions | 2,000 | 100,000 |
| Density (%) | 20 | 6.30 |
| Ratings / user | 20 | 106.04 |
| Ratings / item | 20 | 59.45 |
| Rating mean | 3.52 | 3.39 |
| Rating std. | 1.12 | 0.81 |
| Genres | 5 | 18 |
| Population dispersion $D_{\mathcal{U}}$ | 0.27 | 3.44 |

In the main body of this work we run a series of synthetic experiments to isolate the effects of each element of our theoretical collective action model on the success of a collective engaging in strategic actions. As detailed in above, we use a synthetic construction to simulate the ground truth ratings $R$ which are used to train the model before strategic action. In addition to these synthetic experiments we also reproduce our findings with data drawn from a canonical real movie recommendation dataset, *MovieLens100K* (Harper & Konstan, 2015). For a concrete visualization of these results, please see Figure 7. Here are our findings:

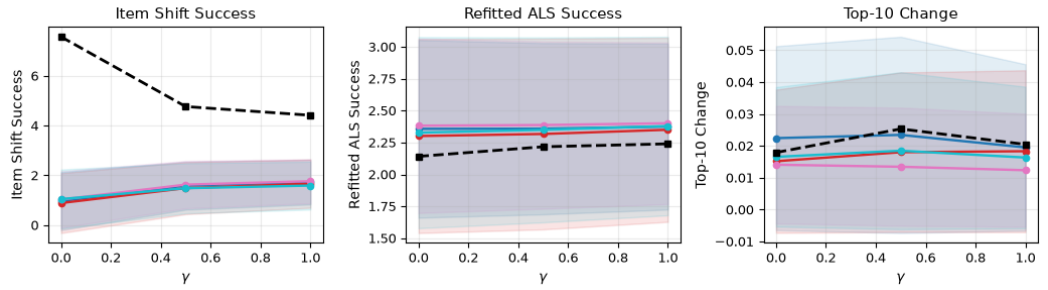
**Larger collectives are consistently more successful, while perfect coordination and cohesion are not necessary for downstream success.** Consistent with our synthetic experiments, increasing the collective proportion $\alpha$ produces substantially greater success after model refitting and generally larger changes in Top-10 exposure. However, the effects of coordination and cohesion are considerably more nuanced. In particular, collectives following our realizable heuristic strategies can achieve greater downstream success while retaining residual heterogeneity or incomplete coordina-



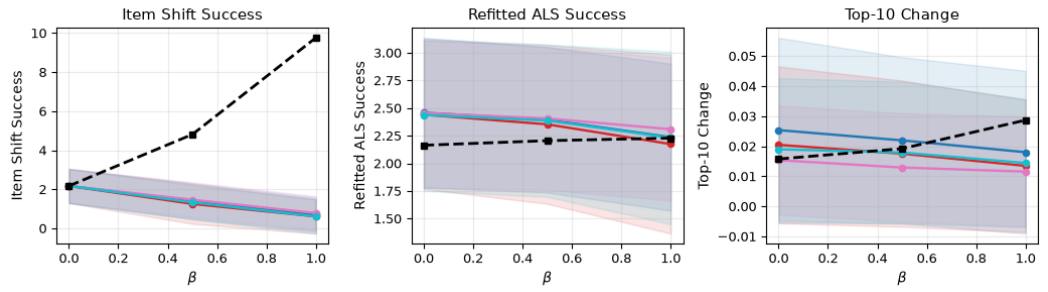

(a) Results for varying collective proportion $\alpha$ under $\beta = 1, tr(\Sigma_C) = \gamma = 0$




(b) Results for varying prior collective heterogeneity $tr(\Sigma_C)$ under $\gamma = 1, \beta = 1, \alpha = 0.1$




(c) Results for varying collective heterogeneity $\gamma$ under $tr(\Sigma_C) = 1, \beta = 1, \alpha = 0.1$




(d) Results for varying collective coordination $\beta$ under $\gamma = 0, \alpha = 0.1$

Figure 7: All results are averaged over 5 independent seeds. For each plot above, we isolate the effects of one parameter, fixing the rest.

tion. Indeed, increasing coordination does not monotonically improve post-refit success for these strategies, and moderate residual heterogeneity can preserve or improve downstream outcomes.

**Optimizing immediate item shift does not necessarily optimize downstream exposure.** This distinction becomes particularly apparent when comparing the realizable heuristic strategies to the bounded oracle. As expected, the oracle substantially outperforms the heuristics on Item-Shift Success, the objective it directly optimizes. Yet this advantage does not consistently persist after the recommender is refit. Across several conditions, heuristic collective actions obtain comparable or greater Post-Refit Success and Top-10 change despite producing substantially smaller immediate shifts in the target item's representation. These results demonstrate that maximizing the immediate representational effect of collective action is not equivalent to maximizing its eventual downstream impact. Instead, the interaction between the collective's action and subsequent model adaptation can favor heterogeneous, imperfectly coordinated behavior.