

# GroupIM: A Mutual Information Maximization Framework for Neural Group Recommendation

Aravind Sankar\*, Yanhong Wu<sup>†</sup>, Yuhang Wu<sup>†</sup>, Wei Zhang<sup>†</sup>, Hao Yang<sup>†</sup>, Hari Sundaram\*

\*University of Illinois at Urbana-Champaign, IL, USA

<sup>†</sup>Visa Research, Palo Alto, CA, USA

{asankar3, hs1}@illinois.edu    †{yanwu, yuhawu, wzhan, haoyang}@visa.com

## ABSTRACT

We study the problem of making item recommendations to ephemeral groups, which comprise users with limited or no historical activities together. Existing studies target persistent groups with substantial activity history, while ephemeral groups lack historical interactions. To overcome group interaction sparsity, we propose data-driven *regularization* strategies to exploit both the *preference covariance* amongst users who are in the same group, as well as the *contextual relevance* of users' individual preferences to each group.

We make two contributions. First, we present a recommender architecture-agnostic framework GroupIM that can integrate arbitrary neural preference encoders and aggregators for ephemeral group recommendation. Second, we regularize the user-group latent space to overcome group interaction sparsity by: maximizing *mutual information* between representations of groups and group members; and dynamically prioritizing the preferences of highly informative members through contextual preference weighting. Our experimental results on several real-world datasets indicate significant performance improvements (31-62% relative NDCG@20) over state-of-the-art group recommendation techniques.

## CCS CONCEPTS

• Information systems → Recommender systems; • Computing methodologies → Neural networks.

## KEYWORDS

Group Recommendation, Neural Collaborative Filtering, Mutual Information, Representation Learning, Data Sparsity

## ACM Reference Format:

Aravind Sankar, Yanhong Wu, Yuhang Wu, Wei Zhang, Hao Yang, Hari Sundaram. 2020. GroupIM: A Mutual Information Maximization Framework for Neural Group Recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '20)*, July 25–30, 2020, Virtual Event, China. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3397271.3401116>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).  
SIGIR '20, July 25–30, 2020, Virtual Event, China

© 2020 Association for Computing Machinery.  
ACM ISBN 978-1-4503-8016-4/20/07...\$15.00  
<https://doi.org/10.1145/3397271.3401116>

## 1 INTRODUCTION

We address the problem of recommending items to *ephemeral* groups, which comprise users who purchase *very few (or no)* items together [26]. The problem is ubiquitous, and appears in a variety of familiar contexts, *e.g.*, dining with strangers, watching movies with new friends, and attending social events. We illustrate key challenges with an example: Alice (who loves Mexican food) is taking a visitor Bob (who loves Italian food) to lunch along with her colleagues, where will they go to lunch? There are three things to note here: first, the group is *ephemeral*, since there is *no* historical interaction observed for this group. Second, *individual preferences may depend on other group members*. In this case, the group may go to a fine-dining Italian restaurant. However, when Alice is with other friends, they may go to Mexican restaurants. Third, groups comprise users with *diverse individual preferences*, and thus the group recommender needs to be cognizant of individual preferences.

Prior work primarily target persistent groups which refer to fixed, stable groups where members have interacted with numerous items *as a group* (*e.g.*, families watching movies). They mainly fall into two categories: heuristic pre-defined aggregation (*e.g.*, least misery [3]) that disregards group interactions; data-driven strategies such as probabilistic models [27, 40] and neural preference aggregators [7, 34]. A key weakness is that these methods either ignore individual user activities [34, 35] or assume that users have the same likelihood to follow individual and collective preferences, across different groups [7, 27, 40]. Lack of expressivity to distinguish the role of individual preferences across groups results in degenerate solutions for sparse ephemeral groups. A few methods exploit external side information in the form of a social network [8, 38], user personality traits and demographics [11], for group decision making. However, side information may often be unavailable.

We train robust ephemeral group recommenders without resorting to any extra side information. Two observations help: first, while groups are ephemeral, group members may have rich individual interaction histories; this can alleviate group interaction sparsity. Second, since groups are ephemeral with sparse training interactions, base group recommenders need reliable guidance to learn informative (non-degenerate) group representations, but the guidance needs to be data-driven, rather than a heuristic.

To overcome group interaction sparsity, our key technical insight is to regularize the latent space of user and group representations in a manner that exploits the *preference covariance* amongst individuals who are in the same group, as well as to incorporate the *contextual* relevance of users' personal preferences to each group.

Thus, we propose two data-driven regularization strategies. *First*, we contrastively regularize the user-group latent space to capture

social user associations and distinctions across groups. We achieve this by maximizing *mutual information* (MI) between representations of groups and group members, which encourages group representations to encode shared group member preferences while regularizing user representations to capture their social associations. *Second*, we contextually identify *informative* group members and regularize the corresponding group representation to reflect their personal preferences. We introduce a novel regularization objective that *contextually* weights users’ personal preferences in each group, in proportion to their user-group MI. *Group-adaptive* preference weighting precludes degenerate solutions that arise during static regularization over ephemeral groups with sparse activities. We summarize our key contributions below:

- **Architecture-agnostic Framework:** To the best of our knowledge, **Group Information Maximization** (GroupIM) is the first recommender architecture-agnostic framework for group recommendation. Unlike prior work [7, 34] that design customized preference aggregators, GroupIM can integrate arbitrary neural preference encoders and aggregators. We show state-of-the-art results with simple efficient aggregators (such as meanpool) that are contrastively regularized within our framework. The effectiveness of meanpool signifies substantially *reduced inference costs* without loss in model expressivity. Thus, GroupIM facilitates straightforward enhancements to base neural recommenders.
- **Group-adaptive Preference Prioritization:** We learn robust estimates of group-specific member relevance. In contrast, prior work incorporate personal preferences through static regularization [7, 27, 40]. We use *Mutual Information* to dynamically learn user and group representations that capture preference covariance across individuals in the same group; and prioritize the preferences of highly relevant members through group-adaptive preference weighting; thus effectively overcoming group interaction sparsity in ephemeral groups. An ablation study confirms the superiority of our MI based regularizers over static alternatives.
- **Robust Experimental Results:** Our experimental results indicate significant performance gains for GroupIM over state-of-the-art group recommenders on four publicly available datasets (relative gains of 31-62% NDCG@20 and 3-28% Recall@20). Significantly, GroupIM achieves stronger gains for: groups of larger sizes; and groups with diverse member preferences.

We organize the rest of the paper as follows. In Section 3, we formally define the problem, introduce a base group recommender unifying existing neural methods, and discuss its limitations. We describe our proposed framework GroupIM in Section 4, present experimental results in Section 5, finally concluding in Section 6.

## 2 RELATED WORK

**Group Recommendation:** This line of work can be divided into two categories based on group types: *persistent* and *ephemeral*. Persistent groups have stable members with rich activity history together, while ephemeral groups comprise users who interact with very few items together [26]. A common approach is to consider persistent groups as virtual users [16], thus, personalized recommenders can be directly applied. However, such methods cannot handle ephemeral groups with sparse interactions. We focus on the more challenging scenario—*recommendations to ephemeral groups*.

Prior work either aggregate recommendation results (or item scores) for each member, or aggregate individual member preferences, towards group predictions. They fall into two classes: *score (or late)* aggregation [3] and *preference (or early)* aggregation [40].

Popular *score aggregation* strategies include least misery [3], average [5], maximum satisfaction [6], and relevance and disagreement [1]. However, these are hand-crafted heuristics that overlook real-world group interactions. Baltrunas et al. [3] compare different strategies to conclude that there is no clear winner, and their relative effectiveness depends on group size and group coherence.

Early *preference aggregation* strategies [39] generate recommendations by constructing a group profile that combines the profiles (raw item histories) of group members. Recent methods adopt a *model-based* perspective to learn data-driven models. *Probabilistic* methods [22, 27, 40] model the group generative process by considering both the personal preferences and relative influence of members, to differentiate their contributions towards group decisions. However, a key weakness is their assumption that users have the same likelihood to follow individual and collective preferences, across different groups. *Neural* methods explore attention mechanisms [2] to learn data-driven preference aggregators [7, 34, 35]. MoSAN [34] models group interactions via sub-attention networks; however, MoSAN operates on persistent groups while ignoring users’ personal activities. AGREE [7] employs attentional networks for joint training over individual and group interactions; yet, the extent of regularization applied on each user (based on personal activities) is the same across groups, which results in degenerate solutions when applied to ephemeral groups with sparse activities.

An alternative approach to tackle interaction sparsity is to exploit external side information, e.g., social network of users [8, 17, 31, 38], personality traits [43], demographics [11], and interpersonal relationships [10, 12]. In contrast, our setting is conservative and does not include extra side information: we know only user and item ids, and item implicit feedback. We address interaction sparsity through novel data-driven regularization and training strategies [19]. Our goal is to enable a wide spectrum of neural group recommenders to seamlessly integrate suitable preference encoders and aggregators.

**Mutual Information:** Recent neural MI estimation methods [4] leverage the InfoMax [21] principle for representation learning. They exploit the *structure* of the input data (e.g., spatial locality in images) via MI maximization objectives, to improve representational quality. Recent advances employ auto-regressive models [24] and aggregation functions [15, 33, 37] with noise-contrastive loss functions to preserve MI between structurally related inputs.

We leverage the InfoMax principle to exploit the preference covariance *structure* shared amongst group members. A key novelty of our approach is MI-guided weighting to regularize group embeddings with the personal preferences of highly relevant members.

## 3 PRELIMINARIES

In this section, we first formally define the ephemeral group recommendation problem. Then, we present a base neural group recommender  $R$  that unifies existing neural methods into a general framework. Finally, we analyze the key shortcomings of  $R$  to discuss motivations for maximizing user-group mutual information.

### 3.1 Problem Definition

We consider the implicit feedback setting (only visits, no explicit ratings) with a user set  $\mathcal{U}$ , an item set  $\mathcal{I}$ , a group set  $\mathcal{G}$ , a binary  $|\mathcal{U}| \times |\mathcal{I}|$  user-item interaction matrix  $X_U$ , and a binary  $|\mathcal{G}| \times |\mathcal{I}|$  group-item interaction matrix  $X_G$ . We denote  $\mathbf{x}_u, \mathbf{x}_g$  as the corresponding rows for user  $u$  and group  $g$  in  $X_U$  and  $X_G$ , with  $|\mathbf{x}_u|, |\mathbf{x}_g|$  indicating their respective number of interacted items.

An *ephemeral group*  $g \in \mathcal{G}$  comprises a set of  $|g|$  users  $u^g = \{u_1^g, \dots, u_{|g|}^g\} \subset \mathcal{U}$  with sparse historical interactions  $\mathbf{x}_g$ .

**Ephemeral Group Recommendation:** We evaluate group recommendation on *strict ephemeral groups*, which have never interacted together before during training. Given a strict ephemeral group  $g$  during testing, our goal is to generate a ranked list over the item set  $\mathcal{I}$  relevant to users in  $u^g$ , i.e., learn a function  $f_G : P(\mathcal{U}) \times \mathcal{I} \mapsto \mathbb{R}$  that maps an ephemeral group and an item to a relevance score, where  $P(\mathcal{U})$  is the power set of  $\mathcal{U}$ .

### 3.2 Base Neural Group Recommender

Several neural group recommenders have achieved impressive results [7, 34]. Despite their diversity in modeling group interactions, we remark that state-of-the-art neural methods share a clear model structure: we present a base group recommender  $R$  that includes three modules: a preference encoder; a preference aggregator; and a joint user and group interaction loss. Unifying these neural group recommenders within a single framework facilitates deeper analysis into their shortcomings in addressing ephemeral groups.

The base group recommender  $R$  first computes user representations  $E \in \mathbb{R}^{|\mathcal{U}| \times D}$  from user-item interactions  $X_U$  using a preference encoder  $f_{\text{ENC}}(\cdot)$ , followed by applying a neural preference aggregator  $f_{\text{AGG}}(\cdot)$  to compute the group representation  $\mathbf{e}_g$  for group  $g$ . Finally, the group representation  $\mathbf{e}_g$  is jointly trained over the group  $X_G$  and user  $X_U$  interactions, to make group recommendations.

**3.2.1 User Preference Representations.** User embeddings  $E$  constitute a latent representation of their personal preferences, indicated in the interaction matrix  $X_U$ . Since latent-factor collaborative filtering methods adopt a variety of strategies (such as matrix factorization, autoencoders, etc.) to learn user embeddings  $E$ , we define the preference encoder  $f_{\text{ENC}} : |\mathcal{U}| \times \mathbb{Z}_2^{|\mathcal{I}|} \mapsto \mathbb{R}^D$  with two inputs: user  $u$  and associated binary personal preference vector  $\mathbf{x}_u$ .

$$\mathbf{e}_u = f_{\text{ENC}}(u, \mathbf{x}_u) \quad \forall u \in \mathcal{U} \quad (1)$$

We can augment  $\mathbf{e}_u$  with additional inputs, including contextual attributes, item relationships, etc. via customized encoders [42].

**3.2.2 Group Preference Aggregation.** A preference aggregator models the interactions among group members to compute an aggregate representation  $\mathbf{e}_g \in \mathbb{R}^D$  for ephemeral group  $g \in \mathcal{G}$ . Since groups are sets of users with no inherent ordering, we consider the class of permutation-invariant functions (such as summation or pooling operations) on sets [41]. Specifically,  $f_{\text{AGG}}(\cdot)$  is permutation-invariant to the order of group member embeddings  $\{e_{u_1}, \dots, e_{u_{|g|}}\}$ . We compute  $\mathbf{e}_g$  using an arbitrary preference aggregator  $f_{\text{AGG}}(\cdot)$  as:

$$\mathbf{e}_g = f_{\text{AGG}}(\{\mathbf{e}_u : u \in u^g\}) \quad \forall g \in \mathcal{G} \quad (2)$$

**3.2.3 Joint User and Group Loss.** The group representation  $\mathbf{e}_g$  is trained over the group-item interactions  $X_G$  with group-loss

$L_G$ . The framework supports different recommendation objectives, including pairwise [29] and pointwise [14] ranking losses. Here, we use a multinomial likelihood formulation owing to its impressive results in user-based neural collaborative filtering [20]. The group representation  $\mathbf{e}_g$  is transformed by a fully connected layer and normalized by a softmax function to produce a probability vector  $\pi(\mathbf{e}_g)$  over  $\mathcal{I}$ . The loss measures the KL-divergence between the normalized purchase history  $\mathbf{x}_g/|\mathbf{x}_g|$  ( $\mathbf{x}_g$  indicates items interacted by group  $g$ ) and predicted item probabilities  $\pi(\mathbf{e}_g)$ , given by:

$$L_G = - \sum_{g \in \mathcal{G}} \frac{1}{|\mathbf{x}_g|} \sum_{i \in \mathcal{I}} x_{gi} \log \pi_i(\mathbf{e}_g); \quad \pi(\mathbf{e}_g) = \text{softmax}(\mathbf{W}_I \mathbf{e}_g) \quad (3)$$

Next, we define the user-loss  $L_U$  that regularizes the user representations  $E$  with user-item interactions  $X_U$ , thus facilitating joint training with shared encoder  $f_{\text{ENC}}(\cdot)$  and predictor ( $\mathbf{W}_I$ ) layers [7]. We use a similar multinomial likelihood-based formation, given by:

$$L_U = - \sum_{u \in \mathcal{U}} \frac{1}{|\mathbf{x}_u|} \sum_{i \in \mathcal{I}} x_{ui} \log \pi_i(\mathbf{e}_u); \quad L_R = L_G + \lambda L_U \quad (4)$$

where  $L_R$  denotes the overall loss of the base recommender  $R$  with balancing hyper-parameter  $\lambda$ . AGREE [7] trains an attentional aggregator with pairwise regression loss over both  $X_U$  and  $X_G$ , while MoSAN [34] trains a collection of sub-attentional aggregators with bayesian personalized ranking [29] loss on just  $X_G$ . Thus, state-of-the-art neural methods AGREE [7] and MoSAN [34] are specific instances of the framework described by base recommender  $R$ .

### 3.3 Motivation

To address ephemeral groups, we focus on *regularization* strategies that are *independent* of the base recommender  $R$ . With the rapid advances in neural methods, we envision future enhancements in neural architectures for user representations and group preference aggregation. Since ephemeral groups by definition purchase very few items together, base recommenders suffer from inadequate training data in group interactions. Here, the group embedding  $\mathbf{e}_g$  receives back-propagation signals from sparse interacted items in  $\mathbf{x}_g$ , thus lacking evidence to reliably estimate the role of each member. To address group interaction sparsity towards robust ephemeral group recommendation, we propose two data-driven regularization strategies that are independent of the base recommendation mechanisms to generate individual and group representations.

**3.3.1 Contrastive Representation Learning.** We note that users' preferences are *group-dependent*; and users occurring together in groups typically exhibit *covarying* preferences (e.g., shared cuisine tastes). Thus, group activities reveal *distinctions* across groups (e.g., close friends versus colleagues) and latent *user associations* (e.g., co-occurrence of users in similar groups), that are not evident when the base recommender  $R$  only predicts sparse group interactions.

We *contrast* the preference representations of group members against those of non-member users with similar item histories, to effectively regularize the latent space of user and group representations. This promotes the representations to encode latent discriminative characteristics shared by group members, that are not discernible from their limited interacted items in  $X_G$ .

**3.3.2 Group-adaptive Preference Prioritization.** To overcome group interaction sparsity, we critically remark that while groups

are ephemeral with sparse interactions, the group members have comparatively richer individual interaction histories. Thus, we propose to *selectively* exploit the personal preferences of group members to enhance the quality of group representations.

The user-loss  $L_U$  (equation 4) in base recommender  $R$  attempts to regularize user embeddings  $E$  based on their individual activities  $X_U$ . A key weakness is that  $L_U$  forces  $e_u$  to *uniformly* predict preferences  $x_u$  across all groups containing user  $u$ . Since groups interact with items differently than individual members, inaccurately utilizing  $X_U$  can become counter-productive. Fixed regularization results in degenerate models that either over-fit or are over-regularized, due to lack of flexibility in adapting preferences per group.

To overcome group interaction sparsity, we *contextually* identify members that are highly *relevant* to the group and regularize the group representation to reflect their personal preferences. To measure contextual relevance, we introduce group-specific relevance weights  $w(u, g)$  for each user  $u$  where  $w(\cdot)$  is a learned weighting function of both user and group representations. This enhances the expressive power of the recommender, thus effectively alleviating the challenges imposed by group interaction sparsity.

In this section, we defined ephemeral group recommendation, and presented a base group recommender architecture with three modules: user representations, group preference aggregation, and joint loss functions. Finally, we motivated the need to: contrastively regularize the user-group space to capture member associations and group distinctions; and learn group-specific weights  $w(u, g)$  to regularize group representations with individual user preferences.

## 4 GROUPIM FRAMEWORK

In this section, we first motivate mutual information towards achieving our two proposed regularization strategies, followed by a detailed description of our proposed framework GROUPIM.

### 4.1 Mutual Information Maximization.

We introduce our user-group mutual information maximization approach through an illustration. We extend the introductory example to illustrate how to regularize Alice’s latent representation based on her interactions in two different groups. Consider Alice who first goes out for lunch to an Italian restaurant with a visitor Bob, and later dines at a Mexican restaurant with her friend Charlie.

First, Alice plays different roles across the two groups (*i.e.*, stronger influence among friends than with Bob) due to the differences in group context (visitors versus friends). Thus, we require a measure to quantify the contextual *informativeness* of user  $u$  in group  $g$ .

Second, we require the embedding of Alice to capture association with both visitor Bob and friend Charlie, yet express variations in her group activities. Thus, it is necessary to not only *differentiate* the role of Alice across groups, but also compute appropriate representations that make her presence in each group more *coherent*.

To achieve these two goals at once, we maximize *user-group mutual information* (MI) to regularize the latent space of user and group representations, and set group-specific relevance weights  $w(u, g)$  in proportion to their estimated MI scores. User-group MI measures the contextual informativeness of a member  $u$  towards the group decision through the reduction in group decision uncertainty when user  $u$  is included in group  $g$ . Unlike correlation measures

that quantify monotonic linear associations, mutual information captures complex non-linear statistical relationships between co-varying random variables. Our proposed MI maximization strategy enables us to achieve our two-fold motivation (Section 3.3):

- **Altering Latent Representation Geometry:** Maximizing user-group MI encourages the group embedding  $e_g$  to encode preference covariance across group members, and regularizes the user embeddings  $E$  to capture social associations in group interactions.
- **Group-specific User Relevance:** By quantifying  $w(u, g)$  through user-group mutual information, we accurately capture the extent of *informativeness* for user  $u$  in group  $g$ , thus guiding group-adaptive personal preference prioritization.

### 4.2 User-Group MI Maximization.

Neural MI estimation [4] has demonstrated feasibility to maximize MI by training a classifier  $D$  (*a.k.a.*, *discriminator* network) to accurately separate *positive* samples drawn from their joint distribution from *negative* samples drawn from the product of marginals.

We maximize user-group MI between group member representations  $\{e_u : u \in u^g\}$  and group representation  $e_g$  (computed in equations 1 and 2 respectively). We train a contrastive *discriminator* network  $D : \mathbb{R}^D \times \mathbb{R}^D \mapsto \mathbb{R}^+$ , where  $D(e_u, e_g)$  represents the probability score assigned to this user-group pair (higher scores for users who are members of group  $g$ ). The positive samples  $(e_u, e_g)$  for  $D$  are the preference representations of  $(u, g)$  pairs such that  $u \in u^g$ , and negative samples are derived by pairing  $e_g$  with the representations of non-member users sampled from a negative sampling distribution  $P_{\mathcal{N}}(u|g)$ . The discriminator  $D$  is trained on a noise-contrastive type objective with a binary cross-entropy (BCE) loss between samples from the joint (positive pairs), and the product of marginals (negative pairs), resulting in the following objective:

$$L_{MI} = -\frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} \frac{1}{\alpha_g} \left[ \sum_{u \in u^g} \log D_{ug} + \sum_{j=1}^{M_g} \mathbb{E}_{\tilde{u} \sim P_{\mathcal{N}}} \log(1 - D_{\tilde{u}g}) \right] \quad (5)$$

where  $\alpha_g = |g| + M_g$ ,  $M_g$  is the number of negative users sampled for group  $g$  and  $D_{ug}$  is a shorthand for  $D(e_u, e_g)$ . This objective maximizes MI between  $e_u$  and  $e_g$  based on the Jensen-Shannon divergence between the joint and the product of marginals [33].

We employ a *preference-biased* negative sampling distribution  $P_{\mathcal{N}}(\tilde{u}|g)$ , which assigns higher likelihoods to *non-member* users who have purchased the group items  $x_g$ . These *hard negative examples* encourage the discriminator to learn latent aspects shared by group members by contrasting against other users with similar individual item histories. We define  $P_{\mathcal{N}}(\tilde{u}|g)$  as:

$$P_{\mathcal{N}}(\tilde{u}|g) \propto \eta \mathcal{I}(x_{\tilde{u}}^T \cdot x_g > 0) + (1 - \eta) \frac{1}{|\mathcal{U}|} \quad (6)$$

where  $\mathcal{I}(\cdot)$  is an indicator function and  $\eta$  controls the sampling bias. We set  $\eta = 0.5$  across all our experiments. In comparison to random negative sampling, our experiments indicate that preference-biased negative user sampling exhibits better discriminative abilities.

When  $L_{MI}$  is trained jointly with the base recommender loss  $L_R$  (equation 4), maximizing user-group MI enhances the quality of user and group representations computed by the encoder  $f_{\text{ENC}}(\cdot)$  and aggregator  $f_{\text{AGG}}(\cdot)$ . We now present our approach to overcome the limitations of the fixed regularizer  $L_U$  (Section 3.3).

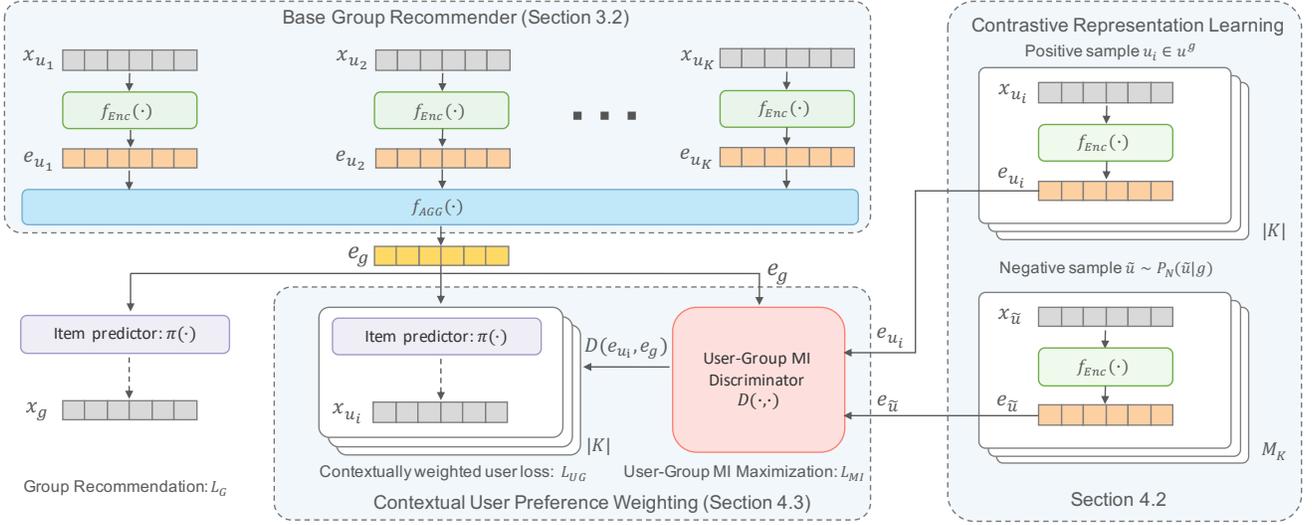


Figure 1: Neural architecture diagram of GroupIM depicting model components and loss terms appearing in Equation 9.

### 4.3 Contextual User Preference Weighting

In this section, we describe a contextual weighting strategy to identify and prioritize personal preferences of relevant group members, to overcome group interaction sparsity. We avoid degenerate solutions by varying the extent of regularization induced by each  $\mathbf{x}_u$  (for user  $u$ ) across groups through group-specific relevance weights  $w(u, g)$ . Contextual weighting accounts for user participation in diverse ephemeral groups with different levels of shared interests.

By maximizing user-group MI, the discriminator  $D$  outputs scores  $D(\mathbf{e}_u, \mathbf{e}_g)$  that quantify the contextual informativeness of each  $(u, g)$  pair (higher scores for informative users). Thus, we set the relevance weight  $w(u, g)$  for group member  $u \in u^g$  to be proportional to  $D(\mathbf{e}_u, \mathbf{e}_g)$ . Instead of regularizing the user representations  $\mathbf{E}$  with  $\mathbf{x}_u$  in each group ( $L_U$  in eqn 4), we directly regularize the group representation  $\mathbf{e}_g$  with  $\mathbf{x}_u$  in proportion to  $D(\mathbf{e}_u, \mathbf{e}_g)$  for each group member  $u$ . Direct optimization of  $\mathbf{e}_g$  (instead of  $\mathbf{e}_u$ ) results in more effective regularization, especially with sparse group activities. We define the contextually weighted user-loss  $L_{UG}$  as:

$$L_{UG} = - \sum_{g \in \mathcal{G}} \frac{1}{|\mathbf{x}_g|} \sum_{i \in \mathcal{I}} \sum_{u \in u^g} D(\mathbf{e}_u, \mathbf{e}_g) x_{ui} \log \pi_i(\mathbf{e}_g) \quad (7)$$

where  $L_{UG}$  effectively regularizes  $\mathbf{e}_g$  with the individual activities of group member  $u$  with contextual weight  $D(\mathbf{e}_u, \mathbf{e}_g)$ .

The overall model objective of our framework GroupIM includes three terms:  $L_G$ ,  $L_{UG}$ , and  $L_{MI}$ , which is described in Section 4.4.4 in detail. GroupIM regularizes the latent representations computed by  $f_{ENC}(\cdot)$  and  $f_{AGG}(\cdot)$  through user-group MI maximization ( $L_{MI}$ ) to contrastively capture group member associations; and contextual MI-guided weighting ( $L_{UG}$ ) to prioritize individual preferences.

### 4.4 Model Details

We now describe the architectural details of preference encoder  $f_{ENC}(\cdot)$ , aggregator  $f_{AGG}(\cdot)$ , discriminator  $D$ , and an alternative optimization approach to train our framework GroupIM.

**4.4.1 User Preference Encoder.** To encode individual user preferences  $X_U$  into preference embeddings  $\mathbf{E}$ , we use a Multi-Layer

Perceptron with two fully connected layers, defined by:

$$\mathbf{e}_u = f_{ENC}(\mathbf{x}_u) = \sigma(\mathbf{W}_2^T (\sigma(\mathbf{W}_1^T \mathbf{x}_u + \mathbf{b}_1) + \mathbf{b}_2))$$

with learnable weight matrices  $\mathbf{W}_1 \in \mathbb{R}^{|\mathcal{I}| \times D}$  and  $\mathbf{W}_2 \in \mathbb{R}^{D \times D}$ , biases  $\mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^D$ , and  $\tanh(\cdot)$  activations for non-linearity  $\sigma$ .

**Pre-training:** We pre-train the weights and biases of the first encoder layer ( $\mathbf{W}_1, \mathbf{b}_1$ ) on the user-item interaction matrix  $X_U$  with user-loss  $L_U$  (equation 4). We use these pre-trained parameters to initialize the first layer of  $f_{ENC}(\cdot)$  before optimizing the overall objective of GroupIM. Our ablation studies in Section 5.5 indicate significant improvements owing to this initialization strategy.

**4.4.2 Group Preference Aggregators.** We consider three preference aggregators MAXPOOL, MEANPOOL and ATTENTION, which are widely used in graph neural networks [13, 30, 36] and have close ties to aggregators examined in prior work, *i.e.*, MAXPOOL and MEANPOOL mirror the heuristics of maximum satisfaction [6] and averaging [5], while attentions learn varying member contributions [7, 34]. We define the three preference aggregators below:

- **Maxpool:** The preference embedding of each member is passed through MLP layers, followed by element-wise max-pooling to aggregate group member representations, given by:

$$\mathbf{e}_g = \text{MAX}(\{\sigma(\mathbf{W}_{AGG} \mathbf{e}_u + \mathbf{b}), \forall u \in u_g\})$$

where MAX denotes the element-wise max operator and  $\sigma(\cdot)$  is a nonlinear activation. Intuitively, the MLP layers compute features for each member, and max-pooling over each of the computed features effectively captures different aspects of group members.

- **Meanpool:** We similarly apply an element-wise mean-pooling operation after the MLP, to compute group representation  $\mathbf{e}_g$  as:

$$\mathbf{e}_g = \text{MEAN}(\{\sigma(\mathbf{W}_{AGG} \mathbf{e}_u + \mathbf{b}), \forall u \in u_g\})$$

- **Attention:** To explicitly differentiate group members' roles, we employ neural attentions [2] to compute a weighted sum of members' preference representations, where the weights are learned by an attention network, parameterized by a single MLP layer.

$$\mathbf{e}_g = \sum_{u \in u_g} \alpha_u \mathbf{W}_{AGG} \mathbf{e}_u \quad \alpha_u = \frac{\exp(\mathbf{h}^T \mathbf{W}_{AGG} \mathbf{e}_u)}{\sum_{u' \in u_g} \exp(\mathbf{h}^T \mathbf{W}_{AGG} \mathbf{e}_{u'})}$$

Dataset	Yelp	Weeplaces	Gowalla	Douban
# Users	7,037	8,643	25,405	23,032
# Items	7,671	25,081	38,306	10,039
# Groups	10,214	22,733	44,565	58,983
# U-I interactions	220,091	1,358,458	1,025,500	1,731,429
# G-I Interactions	11,160	180,229	154,526	93,635
Avg. # items/user	31.3	58.83	40.37	75.17
Avg. # items/group	1.09	2.95	3.47	1.59
Avg. group size	6.8	2.9	2.8	4.2

**Table 1: Summary statistics of four real-world datasets with ephemeral groups. Group-Item interactions are sparse: average number of interacted items per group < 3.5.**

where  $\alpha_u$  indicates the contribution of a user  $u$  towards the group decision. This can be trivially extended to item-conditioned weighting [7], self-attention [35] and sub-attention networks [34].

**4.4.3 Discriminator Architecture.** The discriminator architecture learns a scoring function to assign higher scores to observed  $(u, g)$  pairs relative to negative examples, thus parameterizing group-specific relevance  $w(u, g)$ . Similar to existing work [33], we use a simple bilinear function to score user-group representation pairs.

$$D(\mathbf{e}_u, \mathbf{e}_g) = \sigma(\mathbf{e}_u^T \mathbf{W} \mathbf{e}_g) \quad (8)$$

where  $\mathbf{W}$  is a learnable scoring matrix and  $\sigma$  is the logistic sigmoid non-linearity function to convert raw scores into probabilities of  $(\mathbf{e}_u, \mathbf{e}_g)$  being a positive example. We leave investigation of further architectural variants for the discriminator  $D$  to future work.

**4.4.4 Model Optimization.** The overall objective of GroupIM is composed of three terms, the group-loss  $L_G$  (Equation 3), contextually weighted user-loss  $L_{UG}$  (Equation 7), and MI maximization loss  $L_{MI}$  (Equation 5). The combined objective is given by:

$$L = \underbrace{L_G}_{\text{Group Recommendation Loss}} + \underbrace{\lambda L_{UG}}_{\text{Contextually Weighted User Loss}} + \underbrace{L_{MI}}_{\text{User-Group MI Maximization Loss}} \quad (9)$$

We train GroupIM using an alternating optimization schedule. In the first step, the discriminator  $D$  is held constant, while optimizing the group recommender on  $L_G + \lambda L_{UG}$ . The second step trains  $D$  on  $L_{MI}$ , resulting in gradient updates for both parameters of  $D$  as well as those of the encoder  $f_{\text{ENC}}(\cdot)$  and aggregator  $f_{\text{AGG}}(\cdot)$ .

Thus, the discriminator  $D$  only seeks to *regularize* the model (*i.e.*, encoder and aggregator) during training through loss terms  $L_{MI}$  and  $L_{UG}$ . During inference, we directly use the regularized encoder  $f_{\text{ENC}}(\cdot)$  and aggregator  $f_{\text{AGG}}(\cdot)$  to make group recommendations.

## 5 EXPERIMENTS

In this section, we present an extensive quantitative and qualitative analysis of our model. We first introduce datasets, baselines, and experimental setup (Section 5.1, 5.2, and 5.3), followed by our main group recommendation results (Section 5.4). In Section 5.5, we conduct an ablation study to understand our gains over the base recommender. In Section 5.6, we study how key group characteristics (group size, coherence, and aggregate diversity) impact recommendation results. In Section 5.7, we visualize the variation

in discriminator scores assigned to group members, for different kinds of groups. Finally, we discuss limitations in Section 5.8.

### 5.1 Datasets

First, we conduct experiments on large-scale POI (point-of-interest) recommendation datasets extracted from three location-based social networks. Since the POI datasets do not contain explicit group interactions, we construct group interactions by jointly using the check-ins and social network information: check-ins at the same POI within 15 minutes by an individual and her subset of friends in the social network together constitutes a single group interaction, while the remaining check-ins at the POI correspond to individual interactions. We define the group recommendation task as recommending POIs to ephemeral groups of users. The datasets were pre-processed to retain users and items with five or more check-ins each. We present dataset descriptions below:

- **Weeplaces**<sup>1</sup>: we extract check-ins on POIs over all major cities in the United States, across various categories including Food, Nightlife, Outdoors, Entertainment and Travel.
- **Yelp**<sup>2</sup>: we filter the entire dataset to only include check-ins on restaurants located in the city of Los Angeles.
- **Gowalla** [23]: we use restaurant check-ins across all cities in the United States, in the time period upto June 2011.

Second, we evaluate venue recommendation on Douban, which is the largest online event-based social network in China.

- **Douban** [38]: users organize and participate in social events, where users attend events together in groups and items correspond to event venues. During pre-processing, we filter out users and venues with less than 10 interactions each.

Groups across all datasets are ephemeral since group interactions are sparse (average number of items per group < 3.5 in Table 1)

### 5.2 Baselines

We compare our framework against state-of-the-art baselines that broadly fall into two categories: score aggregation methods with predefined aggregators, and data-driven preference aggregators.

- **Popularity** [9]: recommends items based on item popularity, which is measured by its interaction count in the training set.
- **User-based CF + Score Aggregation**: We utilize a state-of-the-art neural recommendation model VAE-CF [20], followed by score aggregation via: averaging (AVG), least-misery (LM), maximum satisfaction (MAX), and relevance-disagreement (RD).
- **COM** [40]: a probabilistic generative model that considers group members' individual preferences and topic-dependent influence.
- **CrowdRec** [27]: a generative model that extends COM through item-specific latent variables capturing their global popularity.
- **MoSAN** [34]: a neural group recommender that employs a collection of sub-attentional networks to model group member interactions. Since MoSAN originally ignores individual activities  $X_U$ , we include  $X_U$  into  $X_G$  as pseudo-groups with single users.
- **AGREE** [7]: a neural group recommender that utilizes attentional preference aggregation to compute item-specific group member weights, for joint training over individual and group activities.

<sup>1</sup><https://www.yongliu.org/datasets/>

<sup>2</sup><https://www.yelp.com/dataset/challenge>

Dataset Metric	Yelp (LA)				Weeplaces				Gowalla				Douban			
	N@20	N@50	R@20	R@50												
<b>Predefined Score Aggregators</b>																
Popularity [9]	0.000	0.000	0.001	0.001	0.063	0.074	0.126	0.176	0.075	0.088	0.143	0.203	0.003	0.005	0.009	0.018
VAE-CF + AVG [5, 20]	0.142	0.179	0.322	0.513	0.273	0.313	0.502	0.666	0.318	0.362	0.580	0.758	0.179	0.217	0.381	0.558
VAE-CF + LM [3, 20]	0.097	0.120	0.198	0.316	0.277	0.311	0.498	0.640	0.375	0.409	0.610	0.750	0.221	0.252	0.414	0.555
VAE-CF + MAX [6, 20]	0.099	0.133	0.231	0.401	0.229	0.270	0.431	0.604	0.267	0.316	0.498	0.702	0.156	0.194	0.339	0.517
VAE-CF + RD [1, 20]	0.143	0.181	0.321	0.513	0.239	0.279	0.466	0.634	0.294	0.339	0.543	0.723	0.178	0.216	0.379	0.557
<b>Data-driven Preference Aggregators</b>																
COM [40]	0.143	0.154	0.232	0.286	0.329	0.348	0.472	0.557	0.223	0.234	0.326	0.365	0.283	0.288	0.417	0.436
Crowdrec [27]	0.082	0.101	0.217	0.315	0.353	0.370	0.534	0.609	0.325	0.338	0.489	0.548	0.121	0.188	0.375	0.681
AGREE [7]	0.123	0.168	0.332	0.545	0.242	0.292	0.484	0.711	0.160	0.223	0.351	0.605	0.126	0.173	0.310	0.536
MoSAN [34]	0.470	0.494	0.757	<b>0.875</b>	0.287	0.334	0.548	0.738	0.323	0.372	0.584	0.779	0.193	0.239	0.424	0.639
<b>Group Information Maximization Recommenders (GroupIM)</b>																
GroupIM-Maxpool	0.488	0.501	0.676	0.769	0.479	0.505	0.676	0.776	0.433	0.463	0.628	0.747	0.291	0.313	0.524	0.637
GroupIM-Meanpool	0.629	0.637	0.778	0.846	0.518	0.543	0.706	0.804	0.476	0.504	0.682	0.788	0.323	0.351	0.569	0.709
GroupIM-Attention	<b>0.633</b>	<b>0.647</b>	<b>0.782</b>	0.851	<b>0.521</b>	<b>0.546</b>	<b>0.716</b>	<b>0.813</b>	<b>0.477</b>	<b>0.505</b>	<b>0.686</b>	<b>0.796</b>	<b>0.325</b>	<b>0.356</b>	<b>0.575</b>	<b>0.714</b>

Table 2: Group recommendation results on four datasets, R@K and N@K denote the RECALL@K and NDCG@K metrics at  $K = 20$  and 50. The GroupIM variants indicate MAXPOOL, MEANPOOL, and ATTENTION as preference aggregators in our MI maximization framework. GroupIM achieves significant gains of 31 to 62% NDCG@20 and 3 to 28% RECALL@20 over competing group recommenders. Notice that MEANPOOL and ATTENTION variants achieve comparable performance across all datasets.

We tested GroupIM by substituting three preference aggregators, MAXPOOL, MEANPOOL, and ATTENTION (Section 4.4.2). All experiments were conducted on a single Nvidia Tesla V100 GPU with PyTorch [25] implementations on the Linux platform. Our implementation of GroupIM and datasets are publicly available<sup>3</sup>.

### 5.3 Experimental Setup

We randomly split the set of all groups into training (70%), validation (10%), and test (20%) sets, while utilizing the individual interactions of all users for training. Note that each group appears only in one of the three sets. The test set contains *strict ephemeral groups* (i.e., a specific combination of users) that do not occur in the training set. Thus, we train on ephemeral groups and test on strict ephemeral groups. We use NDCG@K and RECALL@K as evaluation metrics.

We tune the latent dimension in the range {32, 64, 128} and other baseline hyper-parameters in ranges centered at author-provided values. In GroupIM, we use two fully connected layers of size 64 each in  $f_{\text{ENC}}(\cdot)$  and tune  $\lambda$  in the range  $\{2^{-4}, 2^{-3}, \dots, 2^6\}$ . We use 5 negatives for each true user-group pair to train the discriminator.

### 5.4 Experimental Results

We note the following key observations from our experimental results comparing GroupIM with its three aggregator variants, against competing baselines on group recommendation (Table 2).

First, heuristic score aggregation with neural recommenders (i.e., VAE-CF) performs comparable to (and often beats) probabilistic models (COM, Crowdrec). Neural methods with multiple non-linear transformations, are expressive enough to identify latent groups of similar users just from their individual interactions.

Second, there is no clear winner among the different pre-defined score aggregation strategies, e.g., VAE-CF + LM (least misery) outperforms the rest on Gowalla and Douban, while VAE-CF + LM (averaging) is superior on Yelp and Weeplaces. This empirically validates the non-existence of a single optimal strategy for all datasets.

Third, MoSAN [34] outperforms both probabilistic models and fixed score aggregators on most datasets. MoSAN achieves better results owing to the expressive power of neural preference aggregators (such as sub-attention networks) to capture group member interactions, albeit not explicitly differentiating personal and group activities. Notice that naive joint training over personal and group activities via static regularization (as in AGREE [7]) results in poor performance due to sparsity in group interactions. Static regularizers on  $X_U$  cannot distinguish the role of users across groups, resulting in models that lack generalisation to ephemeral groups.

GroupIM variants outperform baselines significantly, with ATTENTION achieving overall best results. In contrast to neural methods (i.e., MoSAN and AGREE), GroupIM regularizes the latent representations by contextually weighting the personal preferences of informative members, thus effectively tackling group interaction sparsity. The MAXPOOL variant is noticeably inferior, due to the higher sensitivity of *max* operation to outlier group members.

Note that MEANPOOL performs comparably to ATTENTION. This is because in GroupIM, the discriminator  $D$  does the heavy-lifting of contextually differentiating the role of users across groups to effectively regularize the encoder  $f_{\text{ENC}}(\cdot)$  and aggregator  $f_{\text{AGG}}(\cdot)$  modules. If  $f_{\text{ENC}}(\cdot)$  and  $D$  are expressive enough, efficient MEANPOOL aggregation can achieve near state-of-the-art results (Table 2)

An important implication is the *reduced inference complexity* of our model, i.e., once trained using our MI maximizing framework, simple aggregators (such as meanpool) suffice to achieve state-of-the-art performance. This is especially significant, considering that

<sup>3</sup><https://github.com/CrowdDynamicsLab/GroupIM>

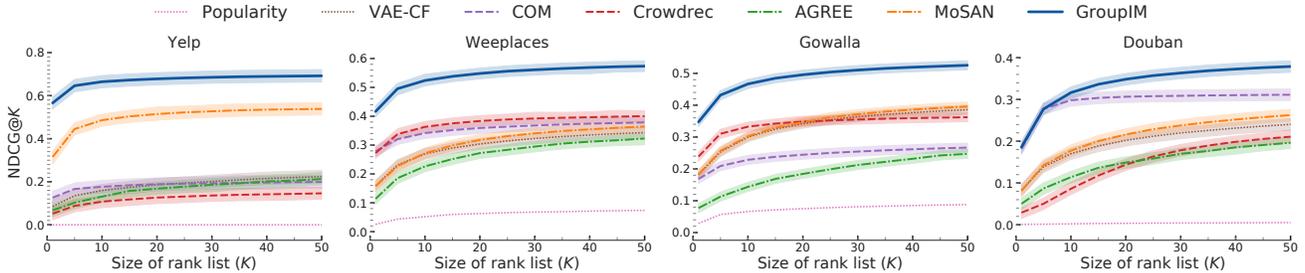


Figure 2: NDCG@K across size of rank list  $K$ . Variance bands indicate 95% confidence intervals over 10 random runs. Existing methods underperform since they either disregard member roles (VAE-CF variants) or overfit to the sparse group activities. GroupIM contextually identifies informative members and regularizes their representations, to show strong gains.

Dataset	Weeplaces		Gowalla	
Metric	N@50	R@50	N@50	R@50
<b>Base Group Recommender Variants</b>				
(1) Base ( $L_G$ )	0.420	0.621	0.369	0.572
(2) Base ( $L_G + \lambda L_U$ )	0.427	0.653	0.401	0.647
<b>GroupIM Variants</b>				
(3) GroupIM ( $L_G + L_{MI}$ )	0.431	0.646	0.391	0.625
(4) GroupIM (Uniform weights)	0.441	0.723	0.418	0.721
(5) GroupIM (Cosine similarity)	0.488	0.757	0.445	0.739
(6) GroupIM (No pre-training)	0.524	0.773	0.472	0.753
(7) GroupIM ( $L_G + \lambda L_{UG} + L_{MI}$ )	<b>0.543</b>	<b>0.804</b>	<b>0.505</b>	<b>0.796</b>

Table 3: GroupIM ablation study (NDCG and RECALL at  $K = 50$ ). Contrastive representation learning (row 3) improves the base recommender (row 1), but is substantially more effective with group-adaptive preference weighting (row 7).

our closest baseline MoSAN [34] utilizes sub-attentional preference aggregation networks that scale quadratically with group size.

We compare the variation in NDCG scores with size of rank list in figure 2. We only depict the best aggregator for VAE-CF. GroupIM consistently generates more *precise* recommendations across all datasets. We observe smaller gains in Douban, where the user-item interactions exhibit substantial correlation with corresponding group activities. GroupIM achieves significant gains in characterizing diverse groups, evidenced by our results in section 5.6.

## 5.5 Model Analysis

In this section, we present an ablation study to analyze several variants of GroupIM, guided by our motivations (Section 3.3). In our experiments, we choose ATTENTION as the aggregator due to its consistently high performance. We conduct studies on Weeplaces and Gowalla to report NDCG@50 and RECALL@50 in Table 3.

First, we examine the base group recommender  $R$  (Section 3.2) which does not utilize MI maximization for model training (Table 3).

**Base Group Recommender.** We examine two variants below:

- (1) We train  $R$  on just group activities  $X_G$  with loss  $L_G$  (equation 3).
- (2) We train  $R$  jointly on individual  $X_U$  and group  $X_G$  activities with static regularization on  $X_U$  using joint loss  $L_R$  (Equation 4).

In comparison to similar neural aggregator MoSAN, our base recommender  $R$  is stronger on NDCG but inferior on RECALL. The difference is likely due to the multinomial likelihood used to train  $R$ , in contrast to the ranking loss in MoSAN. Static regularization

via  $X_U$  (row 1) results in higher gains for Gowalla (richer user-item interactions) with relatively larger margins for RECALL than NDCG.

Next, we examine model variants of GroupIM in two parts:

**GroupIM: Contrastive Representation Learning.** We analyze the benefits derived by just training the contrastive discriminator  $D$  to capture group member associations, *i.e.*, we define a model variant (row 3) to optimize just  $L_G + L_{MI}$ , without the  $L_{UG}$  term. Direct MI maximization (row 3) improves over the base recommender  $R$  (row 1), validating the benefits of contrastive regularization, however still suffers from lack of user preference prioritization.

**GroupIM: Group-adaptive Preference Prioritization.** We analyze the utility of data-driven contextual weighting (via user-group MI), by examining two alternate *fixed* strategies to define  $w(u, g)$ :  
 (4) **Uniform weights:** We assign the same weight  $w(u, g) = 1$  for each group member  $u$  in group  $g$ , when optimizing  $L_{UG}$ .  
 (5) **Cosine similarity:** To model user-group correlation, we set the weight  $w(u, g)$  as the cosine similarity between  $\mathbf{x}_u$  and  $\mathbf{x}_g$ .

From table 3 (rows 4 and 5), the uniform weights variant of loss  $L_{UG}$  (row 4) surpasses the statically regularized model (row 2), due to more direct feedback from  $X_U$  to group embedding  $\mathbf{e}_g$  during model training. Cosine similarity (row 5) achieves stronger gains owing to more accurate correlation-guided user weighting across groups. Our model GroupIM (row 7) has strong gains over the fixed weighting strategies as a result of its regularization strategy to contextually identify informative members across groups.

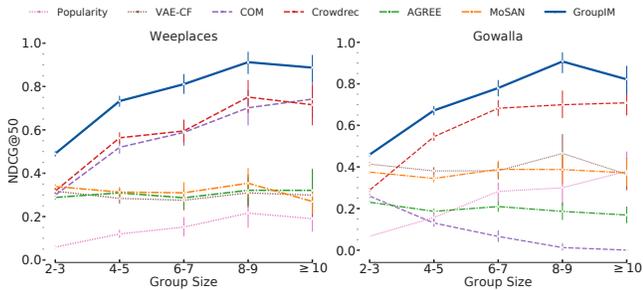
**GroupIM: Pre-training  $f_{ENC}(\cdot)$  on  $X_U$ .** We depict model performance without pre-training (random initializations) in row 6. Our model (row 7) achieves noticeable gains; pre-training identifies good model initialization points for better convergence.

## 5.6 Impact of Group Characteristics

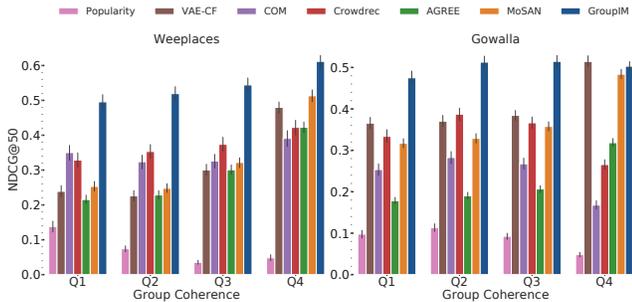
In this section, we examine our results to understand the reason for GroupIM’s gains. We study ephemeral groups along three facets: *group size*; *group coherence*; and *group aggregate diversity*.

**5.6.1 Group Size.** We classify test groups into bins based on five levels of group size (2-3, 4-5, 6-7, 8-9, and  $\geq 10$ ). Figure 3 depicts the variation in NDCG@50 scores on Weeplaces and Gowalla datasets.

We make three key observations: methods that explicitly distinguish individual and group activities (such as COM, CrowdRec, GroupIM), exhibit distinctive trends *wrt* group size. In contrast, MoSAN [34] and AGREE [7], which either uniformly mix both behaviors or apply static regularizers, show no noticeable variation; Performance generally increases with group size. Although test



**Figure 3: Performance (NDCG@50), across group size ranges. GroupIM has larger gains for larger groups due to accurate user associations learnt via MI maximization.**



**Figure 4: Performance (NDCG@50), across group coherence quartiles (Q1: lowest, Q4: highest). GroupIM has larger gains in Q1 & Q2 (low group coherence).**

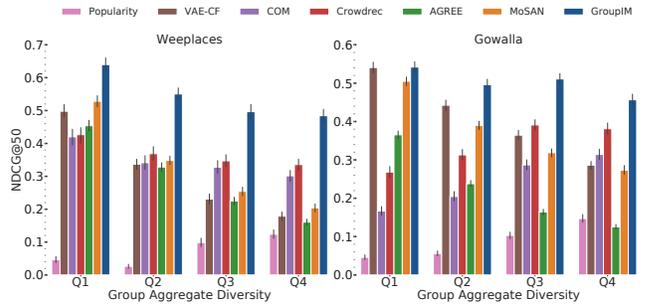
groups are previously unseen, for larger groups, subsets of inter-user interactions are more likely to be seen during training, thus resulting in better performance; **GroupIM** achieves higher (or steady) gains for groups of larger sizes owing to its more accurate prioritization of personal preferences for each member, *e.g.*, **GroupIM** clearly has stronger gains for groups of sizes 8-9 and  $\geq 10$  in Gowalla.

**5.6.2 Group Coherence.** We define *group coherence* as the mean pair-wise correlation of personal activities ( $\mathbf{x}_u$ ) of group members, *i.e.*, if a group has users who frequently co-purchase items, it receives greater coherence. We separate test groups into four quartiles by their coherence scores. Figure 4 depicts NDCG@50 for groups under each quartile (Q1 - Lower values), on Weeplaces and Gowalla.

**GroupIM** has stronger gains for groups with low coherence (quartiles Q1 and Q2), which empirically validates the efficacy of contextual user preference weighting in regularizing the encoder and aggregator, for groups with dissimilar member preferences.

**5.6.3 Group Aggregate Diversity.** We adapt the classical *aggregate diversity* metric [32] to define *group aggregate diversity* as the total number of distinct items interacted across all group members, *i.e.*, if the set of all purchases of group members covers a wider range of items, then the group has higher aggregate diversity. We report NDCG@50 across aggregate diversity quartiles in figure 5.

Model performance typically decays (and stabilizes), with increase in aggregate diversity. Diverse groups with large candidate item sets, pose an information overload for group recommenders, leading to worse results. Contextual prioritization with contrastive learning, benefits diverse groups, as evidenced by the higher relative gains of **GroupIM** for diverse groups (quartiles Q3 and Q4).



**Figure 5: Performance (NDCG@50), across group aggregate diversity quartiles (Q1: lowest, Q4: highest). GroupIM has larger gains in Q3 & Q4 (high diversity).**

### 5.7 Qualitative MI Discriminator Analysis

We examine the contextual weights  $w(u, g)$  estimated by **GroupIM** over test ephemeral groups, across group size and coherence.

We divide groups into four bins based on group sizes (2-3, 4-6, 7-9, and  $\geq 10$ ), and partition them into quartiles based on group coherence within each bin. To analyze the variation in *contextual informativeness* across group members, we compute *MI variation* as the standard deviation of scores given by  $D$  over group members. Figure 6 depicts letter-value plots of *MI variation* for groups in corresponding coherence quartiles across group sizes on Weeplaces.

*MI variation* increases with group size, since larger groups often comprise users with divergent roles and interests. Thus, the discriminator generalizes to unseen groups, to discern and estimate markedly different relevance scores for each group member. To further examine the intuition conveyed by the scores, we compare *MI variation* across group coherence quartiles within each size-range.

*MI variation* is negatively correlated with group coherence for groups of similar sizes, *e.g.*, *MI variation* is consistently higher for groups with low coherence (quartiles Q1 and Q2). For highly coherent groups (quartile Q4),  $D$  assigns comparable scores across all members, which is consistent with our intuitions and earlier results on the efficacy of simple averaging strategies for such groups.

We also analyze parameter sensitivity to user-preference weight  $\lambda$ . Low  $\lambda$  values result in overfitting to the group activities  $X_G$ , while larger values result in degenerate solutions that lack group distinctions (plot excluded for the sake of brevity).

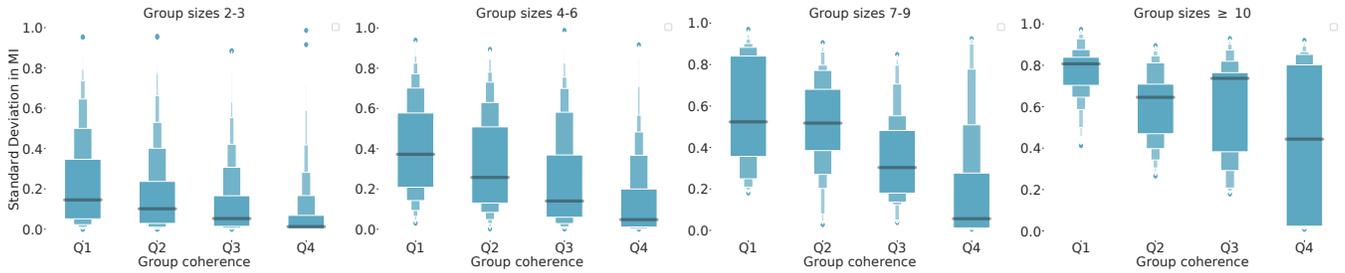
### 5.8 Limitations

We identify two limitations of our work. Despite *learning* to contextually prioritize users' preferences across groups,  $\lambda$  controls the overall strength of preference regularization. Since optimal  $\lambda$  varies across datasets and applications, we plan to explore meta-learning approaches to eliminate such hyper-parameters [28].

**GroupIM** relies on user-group MI estimation to contextually identify informative members, which might become challenging when users have sparse individual interaction histories. In such a scenario, side information (*e.g.*, social network of users), or contextual factors (*e.g.*, location, interaction time) [18] can prove effective.

## 6 CONCLUSION

This paper introduces a recommender architecture-agnostic framework **GroupIM** that integrates arbitrary neural preference encoders



**Figure 6: MI variation (std. deviation in discriminator scores over members) per group coherence quartile across group sizes. For groups of a given size, as coherence increases, MI variation decreases. As groups increase in size, MI variation increases.**

and aggregators for ephemeral group recommendation. To overcome group interaction sparsity, GroupIM regularizes the user-group representation space by maximizing user-group MI to contrastively capture preference covariance among group members. Unlike prior work that incorporate individual preferences through static regularizers, we dynamically prioritize the preferences of informative members through MI-guided contextual preference weighting. Our extensive experiments on four real-world datasets show significant gains for GroupIM over state-of-the-art methods.

## REFERENCES

- [1] Sihem Amer-Yahia, Senjuti Basu Roy, Ashish Chawlat, Gautam Das, and Cong Yu. 2009. Group recommendation: Semantics and efficiency. *VLDB* 2, 1 (2009), 754–765.
- [2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).
- [3] Linas Baltrunas, Tadas Makcinskas, and Francesco Ricci. 2010. Group recommendations with rank aggregation and collaborative filtering. In *RecSys*. 119–126.
- [4] Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeshwar, Sherjil Ozair, Yoshua Bengio, Devon Hjelm, and Aaron Courville. 2018. Mutual Information Neural Estimation. In *ICML*. 530–539.
- [5] Shlomo Berkovsky and Jill Freyne. 2010. Group-based recipe recommendations: analysis of data aggregation strategies. In *RecSys*. ACM, 111–118.
- [6] Ludovico Boratto and Salvatore Carta. 2010. State-of-the-art in group recommendation and new approaches for automatic identification of groups. In *Information retrieval and mining in distributed environments*. Springer, 1–20.
- [7] Da Cao, Xiangnan He, Lianhai Miao, Yahui An, Chao Yang, and Richang Hong. 2018. Attentive group recommendation. In *SIGIR*. ACM, 645–654.
- [8] Da Cao, Xiangnan He, Lianhai Miao, Guangyi Xiao, Hao Chen, and Jiao Xu. 2019. Social-Enhanced Attentive Group Recommendation. *IEEE TKDE* (2019).
- [9] Paolo Cremonesi, Yehuda Koren, and Roberto Turrin. 2010. Performance of recommender algorithms on top-n recommendation tasks. In *RecSys*. 39–46.
- [10] Amra Delic, Judith Masthoff, Julia Neidhardt, and Hannes Werther. 2018. How to Use Social Relationships in Group Recommenders: Empirical Evidence. In *UMAP*. ACM, 121–129.
- [11] Amra Delic, Julia Neidhardt, Thuy Ngoc Nguyen, and Francesco Ricci. 2018. An observational user study for group recommender systems in the tourism domain. *Information Technology & Tourism* 19, 1-4 (2018), 87–116.
- [12] Mike Gartrell, Xinyu Xing, Qin Lv, Aaron Beach, Richard Han, Shivakant Mishra, and Karim Seada. 2010. Enhancing group recommendation by incorporating social relationship interactions. In *GROUPE*. 97–106.
- [13] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *NIPS*. 1024–1034.
- [14] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *WWW*. 173–182.
- [15] R. Devon Hjelm, Alex Fedorov, Samuel Lavoie-Marchildon, Karan Grewal, Philip Bachman, Adam Trischler, and Yoshua Bengio. 2019. Learning deep representations by mutual information estimation and maximization. In *ICLR*.
- [16] Liang Hu, Jian Cao, Guandong Xu, Longbing Cao, Zhiping Gu, and Wei Cao. 2014. Deep modeling of group preferences for group-based recommendation. In *AAAI*.
- [17] Adit Krishnan, Hari Cheruvu, Cheng Tao, and Hari Sundaram. 2019. A modular adversarial approach to social recommendation. In *CIKM*. 1753–1762.
- [18] Adit Krishnan, Mahashweta Das, Mangesh Bendre, Hao Yang, and Hari Sundaram. 2020. Transfer Learning via Contextual Invariants for One-to-Many Cross-Domain Recommendation. *arXiv preprint arXiv:2005.10473* (2020).
- [19] Adit Krishnan, Ashish Sharma, Aravind Sankar, and Hari Sundaram. 2018. An adversarial approach to improve long-tail performance in neural collaborative filtering. In *CIKM*. 1491–1494.
- [20] Dawen Liang, Rahul G Krishnan, Matthew D Hoffman, and Tony Jebara. 2018. Variational autoencoders for collaborative filtering. In *WWW*. 689–698.
- [21] Ralph Linsker. 1988. Self-organization in a perceptual network. *Computer* 21, 3 (1988), 105–117.
- [22] Xingjie Liu, Yuan Tian, Mao Ye, and Wang-Chien Lee. 2012. Exploring personal impact for group recommendation. In *CIKM*. ACM, 674–683.
- [23] Yong Liu, Wei Wei, Aixin Sun, and Chunyan Miao. 2014. Exploiting geographical neighborhood characteristics for location recommendation. In *CIKM*. ACM, 739–748.
- [24] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748* (2018).
- [25] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in PyTorch. In *NIPS-W*.
- [26] Elisa Quintarelli, Emanuele Rabosio, and Letizia Tanca. 2016. Recommending new items to ephemeral groups using contextual user influence. In *RecSys*. ACM, 285–292.
- [27] Vineeth Rakesh, Wang-Chien Lee, and Chandan K Reddy. 2016. Probabilistic group recommendation model for crowdfunding domains. In *WSDM*. ACM, 257–266.
- [28] Mengye Ren, Wenyuan Zeng, Bin Yang, and Raquel Urtasun. 2018. Learning to Reweight Examples for Robust Deep Learning. In *ICML*. 4331–4340.
- [29] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *UAI*. AUAI Press, 452–461.
- [30] Aravind Sankar, Yanhong Wu, Liang Gou, Wei Zhang, and Hao Yang. 2020. DySAT: Deep Neural Representation Learning on Dynamic Graphs via Self-Attention Networks. In *WSDM*. 519–527.
- [31] Aravind Sankar, Xinyang Zhang, Adit Krishnan, and Jiawei Han. 2020. Inf-VAE: A Variational Autoencoder Framework to Integrate Homophily and Influence in Diffusion Prediction. In *WSDM*. 510–518.
- [32] Saul Vargas and Pablo Castells. 2011. Rank and relevance in novelty and diversity metrics for recommender systems. In *RecSys*. ACM, 109–116.
- [33] Petar Veličković, William Fedus, William L. Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. 2019. Deep Graph Infomax. In *ICLR*.
- [34] Lucas Vinh Tran, Tuan-Anh Nguyen Pham, Yi Tay, Yiding Liu, Gao Cong, and Xiaoli Li. 2019. Interact and Decide: Medley of Sub-Attention Networks for Effective Group Recommendation. In *SIGIR*. ACM, 255–264.
- [35] Haiyan Wang, Yuliang Li, and Felix Frimpong. 2019. Group Recommendation via Self-Attention and Collaborative Metric Learning Model. *IEEE Access* 7 (2019), 164844–164855.
- [36] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2019. How Powerful are Graph Neural Networks?. In *ICLR*.
- [37] Yi-Ting Yeh and Yun-Nung Chen. 2019. QAInfomax: Learning Robust Question Answering System by Mutual Information Maximization. In *EMNLP*. 3368–3373.
- [38] Hongzhi Yin, Qinyong Wang, Kai Zheng, Zhixu Li, Jiali Yang, and Xiaofang Zhou. 2019. Social influence-based group representation learning for group recommendation. In *ICDE*. IEEE, 566–577.
- [39] Zhiwen Yu, Xingshe Zhou, Yanbin Hao, and Jianhua Gu. 2006. TV program recommendation for multiple viewers based on user profile merging. *User modeling and user-adapted interaction* 16, 1 (2006), 63–82.
- [40] Quan Yuan, Gao Cong, and Chin-Yew Lin. 2014. COM: a generative model for group recommendation. In *KDD*. ACM, 163–172.
- [41] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan R Salakhutdinov, and Alexander J Smola. 2017. Deep sets. In *NIPS*. 3391–3401.
- [42] Shuai Zhang, Lina Yao, Aixin Sun, and Yi Tay. 2019. Deep Learning Based Recommender System: A Survey and New Perspectives. *ACM Comput. Surv.* 52, 1 (2019), 5:1–5:38.
- [43] Yong Zheng. 2018. Identifying Dominators and Followers in Group Decision Making based on The Personality Traits. In *IUI Workshops*.