Who You Would Like to Share With? A Study of Share Recommendation in Social E-commerce

Houye Ji[1], Junxiong Zhu[2], Xiao Wang[1], Chuan Shi[1*], Bai Wang[1], Xiaoye Tan[2], Yanghua Li[2], Shaojian He[2]

[1]Beijing University of Posts and Telecommunications, Beijing, China
[2]Alibaba Group, Hangzhou, China
{jhy1993,xiaowang,shichuan,wangbai}@bupt.edu.cn,
{xike.zjx,yichen.lyh}@taobao.com,{xiaoye.txy,shaojian.he}@alibaba-inc.com

Abstract

The prosperous development of social e-commerce has spawned diverse recommendation demands, and accompanied a new recommendation paradigm, share recommendation. Significantly different from traditional binary recommendations (e.g., item recommendation and friend recommendation), share recommendation models ternary interactions among $\langle User, Item, Friend\rangle$, which aims to recommend a most likely friend to a user who would like to share a specific item, progressively becoming an indispensable service in social e-commerce. Seamlessly integrating the social relations and purchase behaviours, share recommendation improves user stickiness and monetizes the user influence, meanwhile encountering three unique challenges: rich heterogeneous information, complex ternary interaction, and asymmetric share action. In this paper, we first study the share recommendation problem and propose a heterogeneous graph neural network based share recommendation model, called HGSRec. Specifically, HGSRec delicately designs a tripartite heterogeneous GNNs to describe the multifold characteristics of users and items, and then dynamically fuses them via capturing potential ternary dependency with a dual co-attention mechanism, followed by a transitive triplet representation to depict the asymmetry of share action and predict whether share action happens. Offline experiments demonstrate the superiority of the proposed HGSRec with significant improvements (11.7%-14.5%) over the state-of-the-arts, and online A/B testing on Taobao platform further demonstrates the high industrial practicability and stability of HGSRec.

Introduction

In the era of information explosion, the recommender system has become the most effective way to help users to discover what they are interested in enormous data. As two basic Internet applications, e-commerce and social network both provide the recommender service and therefore generate corresponding item recommendation and friend recommendation, respectively. Recently, with the thriving of online applications, there is a surge of social e-commerce (Gefen and Straub 2004), which integrates social network and e-commerce for better e-commerce service. Benefitting from rich social interactions, social e-commerce provides a new business paradigm which improves user stickiness and activeness and monetizes the user influence. For example, Facebook and Instagram integrate e-commerce into social media, while Amazon and Taobao leverage social interactions to improve e-commerce.

With the development of social e-commerce, a new recommendation paradigm, share recommendation, has sprung up recently. In particular, share recommendation aims to predict whether a user will share an item with his friend. Such recommendation demand is ubiquitous in social e-commerce. Share recommendation has been a unique recommendation paradigm in social e-commerce, due to the following characteristics. Firstly, share recommendation seamlessly integrates the benefits of social relations and item recommendation. Most users coexist in purchase network and social network, so a user well knows his purchasing items, as well as his friends. Seamlessly integrating them, share recommendation not only enhances the stickiness and activeness of users but also monetizes the user influence (e.g., the attention economy and Internet celebrity economy). Secondly, share recommendation provides a reliable recommendation. Since the user knows both the recommended item and his friends, the share action of the user is trustworthy for his friends, which increases the recommendation reliability and thus facilitates purchase action.

The share recommendation is significantly different from traditional recommendations, such as item recommendation (Wang et al. 2019a) and friend recommendation (Wang et al. 2014). As shown in Figure 1, we can find that item recommendation aims to recommend an item to a user (i.e., essentially maximize the probability $P(i_2|u_2)$) and friend recommendation aims to recommend a friend to a user (i.e., maximize the probability $P(u_4|u_2)$). Note that social recommendation (Ma et al. 2008) is naturally item recommendation. Significantly different from the above binary recommendations, the goal of share recommendation is to predict the ternary interactions among $\langle User, Item, Friend\rangle$, i.e., whether a user will share an item with his friend, maximizing the probability $P(u_3|u_2, i_3)$.

Deliberately considering the characteristics of share recommendation, we need to address the following challenges for modeling share recommendation.

- **Rich Heterogeneous Information.** Share recommenda-

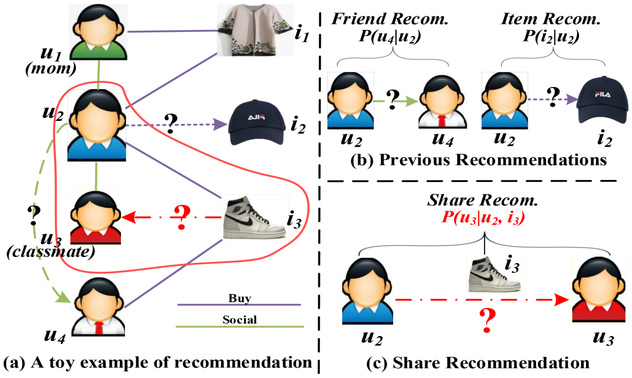


Figure 1: Share recommendation V.S. previous recommendations.

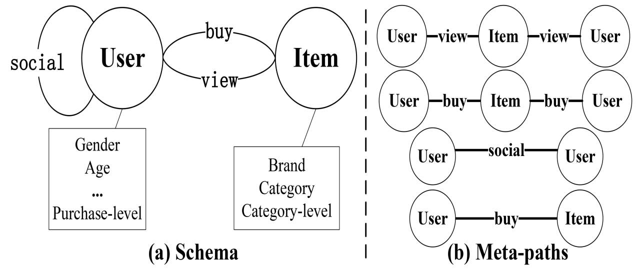


Figure 2: A typical example for share recommendation.

tion usually contains complex heterogeneous information, including complex interactions among users and items, as well as rich feature information of users and items. Such an example is shown in Figure 2(a). How to handle the complex interactions and utilize the diverse features simultaneously is an urgent problem that needs to be solved.

• **Complex ternary interaction.** Different from simple binary interaction in traditional recommendations, exemplified as $\langle u_2, i_2 \rangle$ interaction in the item recommendation and $\langle u_2, u_4 \rangle$ interaction in friend recommendation in Figure 1, share recommendation faces complex ternary interaction (e.g., $\langle u_2, i_3, u_3 \rangle$ in Figure 1). We need to consider the suitability of a share action, which evaluates the matching degree of three objects (e.g., u_2, i_3, u_3) in the share action. According to the characteristic of the recommended item, a user will recommend it to an appropriate friend, and thus how the item influence the user (or the friend) should be considered. Taking Figure 1 as an example, the user u_2 will share the shoes i_3 to his classmate u_3, rather than his mom u_1. So we need to model the ternary interaction of user, item, and friend, considering their suitability.

• **Asymmetric Share Action.** The share action is asymmetric and irreversible, which means the share action may not happen if we swap the roles of the user and the friend. As shown in Figure 1, the user u_2 may share a women overcoat i_1 to his mom u_1, while the user u_1 would not share the women overcoat i_1 to her son u_2. Therefore, a desired model should consider the asymmetry of share action.

In this paper, we first study the problem of share recommendation and propose a novel **H**eterogeneous **G**raph neural network based **S**hare **Rec**ommendation model (**HGSRec**). We model the share recommendation system as an attributed heterogeneous graph to integrate rich heterogeneous information, and then we design HGSRec to learn the embeddings of u, i, v and predict the probability of share action $\langle u, i, v \rangle$ happening. Specifically, after initializing node embedding via encoding rich node features, a tripartite heterogeneous GNNs is designed to learn the embeddings of u, i, v, respectively, via aggregating their meta-path based neighbors, which enables HGSRec flexibly fuse different aspects of information. Furthermore, a dual co-attention mechanism is proposed to dynamically fuse the multiple embeddings of u (or v) under different meta-paths, considering the influence of item i to user u (or v), to improve the suitability of $\langle u, i, v \rangle$. Finally, a transitive triplet representation of $\langle u, i, v \rangle$ is employed to predict whether share action happens.

The contributions of our work are summarized as follows:

• We study a newly emerging recommendation problem, share recommendation, which aims to predict whether a user will recommend an item to his friend. Different from traditional binary recommendations, the share recommendation provides a ternary recommendation paradigm.

• We propose a novel HGSRec for share recommendation, with the help of delicate designs, such as tripartite heterogeneous GNNs, dual co-attention mechanism and transitive triplet representation.

• Extensive experiments on Taobao demonstrate the superiority of the proposed HGSRec with more than 10% performance improvement, compared to the state-of-the-arts.

Related Work

Item recommendation (Sarwar et al. 2001; Shi et al. 2018; Hu et al. 2018), which aims to predict whether one user will buy or view one item, has been extensively studied and provided great economic value. Several works (Ma et al. 2008; Fan et al. 2019b) leverage social information to further improve the performance of item recommendation. People recommendation (Kutty, Nayak, and Chen 2014; Ricci, Rokach, and Shapira 2011) aims to predict whether one user will interact with another user, such as user-friend in friend recommendation, employer-employee in job recommendation, and male-female in dating recommendation. Essentially, both item recommendation and people recommendation consider the binary interaction, such as $\langle User, Item \rangle$ and $\langle User, User \rangle$. Significantly different from the above recommendations, share recommendation focus on ternary interaction $\langle User, Item, Friend \rangle$ and predict whether one user will share one item with his friend.

Graph neural networks (Kipf and Welling 2017; Veličković et al. 2018; Hamilton, Ying, and Leskovec 2017) generalize deep learning to graph-structured data, which usually follows the message-passing framework to receive messages from neighbors and apply neural network to update node embedding. Kipf et al. (Kipf and Welling 2017) propose graph convolutional network for node classification, and (Hamilton, Ying, and Leskovec 2017; Veličković et al. 2018) propose diverse aggregating functions. Several works

(Ying et al. 2018; Fan et al. 2019b; Wang et al. 2019a; Wu et al. 2019) generalize GNNs to perform item recommendation. Recently, some works (Wang et al. 2019b; Fan et al. 2019a; Hu et al. 2019; Wang et al. 2020) extend GNNs for heterogeneous graph and generalize them (Fan et al. 2019a; Zhao et al. 2019) for recommendation. However, all previous GNN based recommendation methods focus on binary interaction and cannot be applied to model ternary interaction $\langle User, Item, Friend\rangle$ in share recommendation.

Preliminaries

Definition 1. ***Attributed Heterogeneous Graph****. An attributed heterogeneous graph, denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$, where $\mathcal{V} = \mathcal{V}_U \cup \mathcal{V}_I$ is the node sets, $\mathcal{E} = \mathcal{E}_S \cup \mathcal{E}_O$ is the edge sets, $\mathbf{X} \in \mathbb{R}^{|\mathcal{V}|\times K}$ is an attribute matrix of nodes. Here $\mathcal{V}_U$ and $\mathcal{V}_I$ are the sets of users and items, respectively. $\mathcal{E}_S = \langle \mathcal{V}_U, \mathcal{V}_U\rangle$ denotes $User$-$User$ interaction and $\mathcal{E}_O = \langle \mathcal{V}_U, \mathcal{V}_I\rangle$ denotes $User$-$Item$ interaction. For $u, v \in \mathcal{V}_U$, v is u's* ***friend*** *if $\langle u, v\rangle \in \mathcal{E}_S$ and the friend set of u is $\mathcal{F}(u) = \{v|\langle u, v\rangle \in \mathcal{E}_S\}$.*

Example. Figure 2(a) shows the attributed heterogeneous graph of share recommendation. Here u_2 has two friends denoting as $\mathcal{F}(u_2) = \{u_1, u_3\}$. Meta-path (Sun et al. 2011), a composite relation connecting two nodes, is able to extract rich semantics. As shown in Figure 2(b), $User \xrightarrow{buy} Item \xrightarrow{buy} User$ (*U-b-I-b-U* for short) meaning the co-buying relations, $User \xrightarrow{social} User$ (*U-s-U* for short) meaning the social relations, $User \xrightarrow{buy} Item$ (*U-b-I* for short) meaning buy relations, and $User \xrightarrow{view} Item \xrightarrow{view} User$ (*U-v-I-v-U* for short) meaning the co-viewing relations.

Definition 2. ***Share Recommendation****. Given an attributed heterogeneous graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$ representing a share recommendation system, share recommendation aims to predict a share action $\langle u, i, v\rangle$ (formulated with $\langle User, Item, Friend\rangle$, or abbreviated with $\langle U, I, V\rangle$). Specifically, the purpose of share recommendation is to recommend the most likely $Friend$ $v \in \mathcal{F}(u)$ to $User$ $u \in \mathcal{V}_U$ who would like to share the $Item$ $i \in \mathcal{V}_I$ ($\langle u, i\rangle \in \mathcal{E}_O$), i.e., $\arg\max_v P(v|u, i)$. The label $y_{u,i,v} \in \{0, 1\}$ indicates whether share action happens.*

Example. As shown in Figure 1(c), share recommendation will recommend a most likely friend, like $u_3 \in \mathcal{F}(u_2)$, to a user u_2 who would like to share the shoes i_3, which essentially maximizes the probability $P(u_3|u_2, i_3)$.

The Proposed Model

In this section, we present a novel **H**eterogeneous **G**raph neural network based **S**hare **Rec**ommendation (**HGSRec**). The overall framework of HGSRec is shown in Figure 3.

Initialization with Feature Embedding

Firstly, we initialize node embedding via embedding their features. Different from ID embedding, feature embedding has two-fold benefits: (1) In real applications, there are numerous of newly coming nodes every day. The feature embedding effectively generates embeddings for previously unseen nodes by utilizing their features. (2) The number of features is much less than the number of nodes, which significantly reduces the number of learnable parameters.

For the k-th node feature $f_k \in \mathbb{R}^{|f_k|*1}$, we initialize a feature embedding matrix $\mathbf{M}^{f_k} \in \mathbb{R}^{d*|f_k|}$, where $|f_k|$ means the number of values of feature f_k and d is the dimension of feature embedding. The embedding of u's k-th feature is shown as follows:

$$\mathbf{e}_u^{f_k^U} = \mathbf{M}^{f_k^U} \cdot u^{f_k^U}. \tag{1}$$

Considering all the features of user u, we can get the initial user embedding $\mathbf{x}_u$, as follows:

$$\mathbf{x}_u = \sigma\left(\mathbf{W}_U \cdot \left(\overset{|f^U|}{\underset{k=1}{\|}} \mathbf{e}_u^{f_k^U}\right) + \mathbf{b}_U\right), \tag{2}$$

where $||$ denotes the concatenate operation, $\mathbf{W}_U$ and $\mathbf{b}_U$ denote the weight matrix and bias vector, respectively. The same process can be done for item/friend embedding.

Tripartite Heterogeneous Graph Neural Networks

Here we propose tripartite heterogeneous GNNs to learn embeddings of u, i, v via corresponding heterogeneous GNN (i.e., $HeteGNN^U$, $HeteGNN^I$, and $HeteGNN^V$), respectively. Heterogeneous GNN usually follows a hierarchical manner: It first aggregates information from one kind of neighbors via one meta-path and learns the semantic-specific node embeddings in node-level. Then, it aggregates multiple semantics from different meta-paths and fuses a set of semantic-specific node embeddings in semantic-level.

Specifically, given one user u and k_1 user-related meta-paths $\{\Phi_1^U, \Phi_2^U, \cdots, \Phi_{k_1}^U\}$, $HeteGNN^U$ is able to get k_1 semantic-specific user embeddings $\{\mathbf{x}_u^{\Phi_1^U}, \mathbf{x}_u^{\Phi_2^U}, \cdots, \mathbf{x}_u^{\Phi_{k_1}^U}\}$.

$$\mathbf{x}_u^{\Phi_1^U}, \mathbf{x}_u^{\Phi_2^U}, \cdots, \mathbf{x}_u^{\Phi_{k_1}^U} = HeteGNN^U(u). \tag{3}$$

Note that the number of meta-path based neighbors of different nodes could be quite different, so we need to sample fixed number of neighbors. Random sampling strategy causes heavy computation consumption and missing important nodes. Here we propose a top-N semantic sampling strategy: (1) If the number of meta-path based neighbors is more than fixed number N, we sample top-N meta-path based neighbors based on connection strength (e.g., how many times a user view an item). (2) Or else, we adopt resample to get N meta-path based neighbors.

Given a user u and corresponding meta-path Φ^U, we propose a novel semantic aggregator $SemAgg_u^{\Phi^U}$ to aggregate sampled neighbors $\mathcal{N}_u^{\Phi^U}$ and obtain the meta-path based embedding $\mathbf{x}_u^{\mathcal{N}_u^{\Phi^U}}$, as follows:

$$\mathbf{x}_u^{\mathcal{N}_u^{\Phi^U}} = SemAgg_u^{\Phi^U}(\{\mathbf{x}_n|\forall n \in \mathcal{N}_u^{\Phi^U}\}). \tag{4}$$

Considering the time efficiency, we adopt $MeanPooling$ to accelerate aggregating processing for faster prediction. The semantic aggregator $SemAgg_u^{\Phi^U}$ is shown as follows:

$$\mathbf{x}_u^{\mathcal{N}_u^{\Phi^U}} = MeanPooling(\{\mathbf{x}_n|\forall n \in \mathcal{N}_u^{\Phi^U}\}). \tag{5}$$

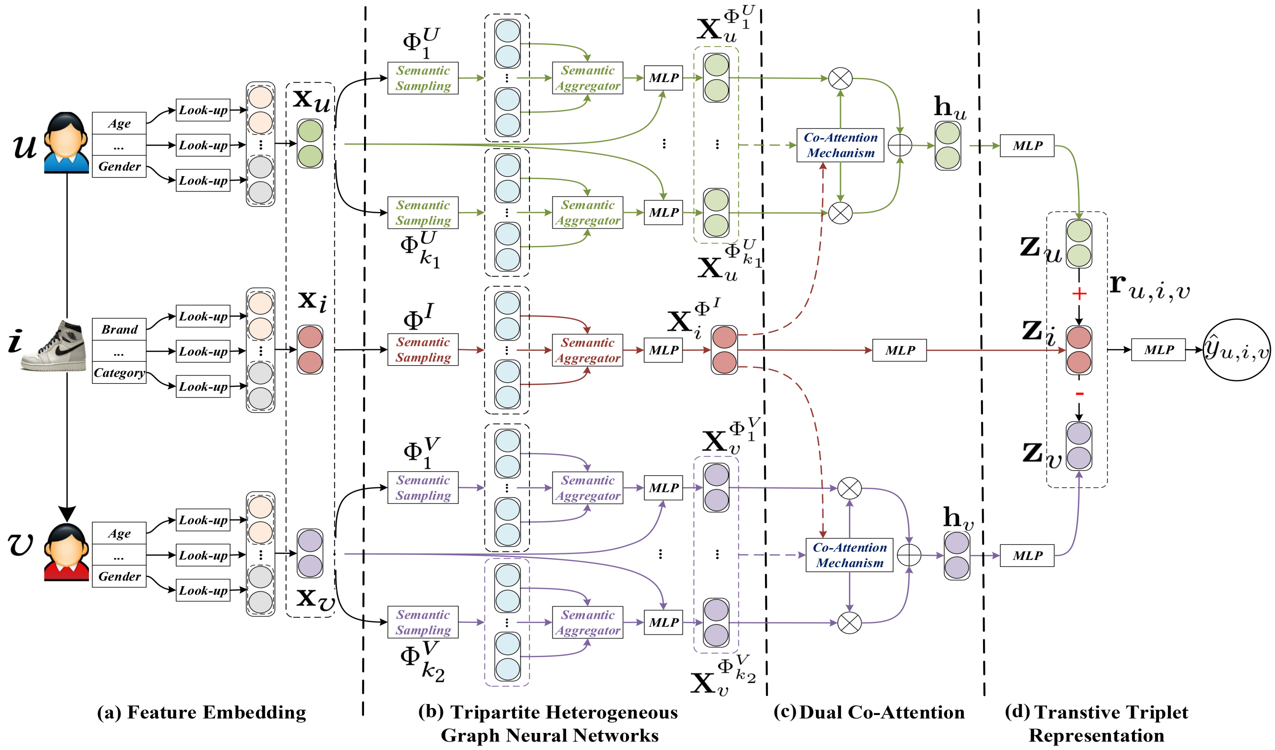


Figure 3: The overall framework of the proposed HGSRec. (a) Initializing user and item embedding via feature embedding. (b) Updating node embedding via tripartite heterogeneous graph neural networks. (c) Fusing embedding dynamicially via the dual co-attention mechanism. (d) Modeling asymmetric share action via transitive triplet representation.

To emphasize the property of user u explicitly, we concatenate initial embedding $\mathbf{x}_u$ and meta-path based embedding $\mathbf{x}_u^{\mathcal{N}_u^{\Phi^U}}$ and get the semantic-specific user embedding $\mathbf{x}_u^{\Phi^U}$,

$$\mathbf{x}_u^{\Phi^U} = \sigma(\mathbf{W}^{\Phi^U} \cdot (\mathbf{x}_u || \mathbf{x}_u^{\mathcal{N}_u^{\Phi^U}}) + \mathbf{b}^{\Phi^U}), \tag{6}$$

where $\mathbf{W}^{\Phi^U}$ and $\mathbf{b}^{\Phi^U}$ denote the weight matrix and bias vector for meta-path Φ^U, respectively. Given a set of user-related meta-paths $\{\Phi_1^U, \Phi_2^U, \cdots, \Phi_{k_1}^U\}$, we can get k_1 semantic-specific user embeddings $\{\mathbf{x}_u^{\Phi_1^U}, \mathbf{x}_u^{\Phi_2^U}, \cdots, \mathbf{x}_u^{\Phi_{k_1}^U}\}$ which describe the characteristics of user u from different aspects. The same process can be done via $HeteGNN^V$ to learn multiple semantic-specific embeddings $\{\mathbf{x}_v^{\Phi_1^V}, \mathbf{x}_v^{\Phi_2^V}, \cdots, \mathbf{x}_v^{\Phi_{k_2}^V}\}$ of friend v. Since the characteristic of the item is much simple and stable than the user, we only adopt one meta-path Φ^I to get the embedding $\mathbf{x}_i^{\Phi^I}$ of item i via $HeteGNN^I$.

Dual Co-Attention Mechanism

After obtaining a set of semantic-specific node embeddings (e.g., $\{\mathbf{x}_u^{\Phi_1^U}, \mathbf{x}_u^{\Phi_2^U}, \cdots, \mathbf{x}_u^{\Phi_{k_1}^U}\}$), we aim to fuse them properly based on the complex ternary interactions $\langle u, i, v\rangle$. So a dual co-attention mechanism is designed to dynamically fuse the embeddings of u (or v) under different meta-paths, considering the effect of item i. which consist of co-attention mechanism $CoAtt_{U,I}$ for $\langle U, I\rangle$ and co-attention mechanism $CoAtt_{V,I}$ for $\langle V, I\rangle$. Specifically, it learns the interaction-specific attention values of meta-paths for $\langle u, i, v\rangle$ and get the most appropriate embedding of u, v, with the following benefits: (1) It reinforces the dependency of $\langle u, i, v\rangle$, making HGSRec more integrated. (2) It dynamically fuses the embeddings of u (or v), improving share suitabilities.

Taking $\langle U, I\rangle$ as an example, the co-attention mechanism $CoAtt_{U,I}$ aims to learn a set of interaction-specific co-attention weights $\{w_{u,i}^{\Phi_1^U}, w_{u,i}^{\Phi_2^U}, \cdots, w_{u,i}^{\Phi_{k_1}^U}\}$ for user u,

$$w_{u,i}^{\Phi_1^U}, w_{u,i}^{\Phi_2^U}, \cdots, w_{u,i}^{\Phi_{k_1}^U} = \text{CoAtt}_{U,I}(\mathbf{x}_u^{\Phi_1^U}, \cdots, \mathbf{x}_u^{\Phi_{k_1}^U}, \mathbf{x}_i^{\Phi^I}). \tag{7}$$

Specifically, we concatenate the semantic-specific embedding of u and i and project them into co-attention space. Then, we adopt a co-attention vector $\mathbf{q}_{U,I}$ to learn the importances of meta-paths for user u. The importance of meta-path Φ_m^U for u in the interaction $\langle u, i\rangle$, denoted as $\alpha_{u,i}^{\Phi_m^U}$,

$$\alpha_{u,i}^{\Phi_m^U} = \mathbf{q}_{U,I}^T \cdot \sigma(\mathbf{W}^{U,I} \cdot (\mathbf{x}_u^{\Phi_m^U} || \mathbf{x}_i^{\Phi^I}) + \mathbf{b}^{U,I}), \tag{8}$$

where $\mathbf{W}^{U,I}$ and $\mathbf{b}^{U,I}$ denote the weight matrix and bias vector, respectively. After obtaining the importances of meta-paths, we normalize them via softmax to get the co-attention weight $w_{u,i}^{\Phi_m^U}$ of meta-path Φ_m^U, shown as follows:

$$w_{u,i}^{\Phi_m^U} = \frac{\exp(\alpha_{u,i}^{\Phi_m^U})}{\sum_{m=1}^{k_1} \exp(\alpha_{u,i}^{\Phi_m^U})}, \tag{9}$$

where $w_{u,i}^{\Phi_m^U}$ reflects the contribution of meta-path Φ_m^U in improving share suitability. Larger $w_{u,i}^{\Phi_m^U}$ means the meta-path Φ_m^U of u is more suitable to the item i which makes higher contribution in improving the suitability of $\langle u, i, v\rangle$. With the learned weights as coefficients, we can obtain the fused embedding $\mathbf{h}_u$ of u, shown as follows:

$$\mathbf{h}_u = \sum_{m=1}^{k_1} w_{u,i}^{\Phi_m^U} \cdot \mathbf{x}_u^{\Phi_m^U}. \tag{10}$$

Obviously, $\mathbf{h}_u$ is dynamically changed with regard to different co-attention weights, where the co-attention weights are dynamically changed with regard to different items.

Similar to $CoAtt_{U,I}$, $CoAtt_{V,I}$ learns a set of co-attention weights $\{w_{v,i}^{\Phi_1^V}, w_{v,i}^{\Phi_2^V}, \cdots, w_{v,i}^{\Phi_{k_2}^V}\}$ for friend v and get the fused friend embeddings $\mathbf{h}_v$. Since we only select one meta-path for item, the fused embedding $\mathbf{h}_i$ of item i is actually $\mathbf{x}_i^{\Phi^I}$.

Transitive Triplet Representation

To predict the share action $\langle u, i, v\rangle$, we need to construct a triplet representation $\mathbf{r}_{u,i,v}$ based on $\mathbf{h}_u, \mathbf{h}_i, \mathbf{h}_v$. We first project all types of nodes in $\langle U, I, V\rangle$ into the same space via three type-specific MLPs, shown as follows:

$$\mathbf{z}_u = MLP^U(\mathbf{h}_u),\ \mathbf{z}_i = MLP^I(\mathbf{h}_i),\ \mathbf{z}_v = MLP^V(\mathbf{h}_v). \tag{11}$$

A simple way to construct the triplet representation $\mathbf{r}_{u,i,v}$ is to concatenate all node embeddings (a.k.a., $\mathbf{z}_u||\mathbf{z}_i||\mathbf{z}_v$). However, the simple concatenation cannot explicitly capture the remarkable characteristics of share action: (1) The share recommendation actually aims to rank candidate friends based on both user and item (e.g., calculate the similarity between $\mathbf{z}_u + \mathbf{z}_i$ and $\mathbf{z}_v$), so the share action is asymmetric and the roles of user and friend cannot be exchanged. (2) The item describes the transition between user and friend, so it is an indispensable bridge in establishing share action.

Inspired by relational translation (Antoine et al. 2013), we propose a transitive triplet representation $\mathbf{r}_{u,i,v}$ to explicitly model the characteristics of share action via item-translating, shown as follows:

$$\mathbf{r}_{u,i,v} = |\mathbf{z}_u + \mathbf{z}_i - \mathbf{z}_v|, \tag{12}$$

where $|\cdot|$ denotes the absolute operation. Then, we feed $\mathbf{r}_{u,i,v}$ into MLP and get the predict score $\hat{y}_{u,i,v}$, as follows:

$$\hat{y}_{u,i,v} = \sigma(\mathbf{W} \cdot \mathbf{r}_{u,i,v} + b), \tag{13}$$

where $\mathbf{W}$ and b denote the weight vector and bias scalar, respectively. Finally, we calculate cross-entropy loss,

$$L = \sum_{u,i,v \in \mathcal{D}} (y_{u,i,v} \log \hat{y}_{u,i,v} + (1 - y_{u,i,v}) \log(1 - \hat{y}_{u,i,v})), \tag{14}$$

where $y_{u,i,v}$ is the label of the triplet, $\mathcal{D}$ denotes the dataset. Then, we analyze the space complexity of HGSRec. The learnable parameters in HGSRec mainly come from embedding matrixes rather than neural networks. Assuming we have A node IDs and B node features, the number of parameters of ID embedding and feature embedding are $O(A * d)$ and $O(B * d)$, respectively. In the share scenario, the number of IDs A (usually billion-level) is significantly more than the number of features B so as to $O(A * d) \gg O(B * d)$, which means HGSRec is able to efficiently handle large-scale data.

Dataset	3-days	4-days	5-days
#Train $\langle u, i, v\rangle$	3,324,367	4,443,996	5,611,531
#Train $Users$	1,064,426	1,315,126	1,546,017
#Train $Items$	537,048	679,784	818,290
#Valid $\langle u, i, v\rangle$	1,401,395		
#Valid $Users$	539,959		
#Valid $Items$	247,907		

Table 1: The statistics of the datasets.

Experiments

Datasets We collect data from Taobao platform, ranging from 2019/10/09 to 2019/10/14, and construct an attributed heterogeneous graph (shown in Figure 2). Each sample contains a share action $\langle u, i, v\rangle$ and corresponding label $y_{u,i,v} \in \{0, 1\}$. We select four meta-paths including *U-s-U*, *U-b-I-b-U* and *U-v-I-v-U* for the user and *U-b-I* for the item. In offline experiments, we use the last day (i.e., 2019/10/14) as validation set and the previous 3/4/5 days as training sets, marked as **3-days**, **4-days**, and **5-days**, respectively. To comprehensively evaluate the results, we vary the size of each training set from 40% to 100%. The details of the datasets are shown in Table 1.

Baselines We select feature based models (i.e., LR, DNN, and XGBoost) and GNN models (i.e., GraphSAGE, IGC, and MEIRec) as baselines. Since IGC and MEIRec cannot handle ternary recommendation, we also provide tripartite versions (i.e., IGC+ and MEIRec+) for share recommendation. To validate delicate designs in HGSRec, we also test two variants of HGSRec (HGSRec$_{\backslash att}$ and HGSRec$_{\backslash tra}$). Note that although deep models depend on randomness whose performances change with different random seeds, their performances on large-scale Taobao datasets are quite stable ($a.k.a$, the variance of HGSRec less than 0.001).

• **LR/DNN/XGBoost**: They are classical algorithms for industry. We concatenate the feature of user, item and friend as the model input and predict the share action $\langle u, i, v\rangle$.

• **GraphSAGE** (SAGE for short) (Hamilton, Ying, and Leskovec 2017): It is a classical GNN which leverages sampler and aggregator to embed homogeneous graph. Since GraphSAGE cannot perform share recommendation directly, we ignore the influence of item and utilize it to perform people-to-people recommendation ($a.k.a$, $\langle u, v\rangle$).

• **IGC/IGC+** (Zhao et al. 2019): It is a HeteGNN based recommendation model. Since IGC cannot perform share recommendation directly, we first recall all candidate friends of user u and then predict the $\langle i, v\rangle$. Then, we extend IGC as IGC+ to learn the embedding of u, i, v and concatenate them to predict $\langle u, i, v\rangle$.

• **MEIRec/MEIRec+** (Fan et al. 2019a): It is a HeteGNN

Model	3-days				4-days				5-days			
	40%	60%	80%	100%	40%	60%	80%	100%	40%	60%	80%	100%
LR	67.56	67.62	67.26	67.69	67.58	67.65	67.68	67.72	67.62	67.67	67.72	67.74
XGBoost	72.04	72.14	72.13	72.18	72.08	72.11	72.15	72.49	72.72	72.54	71.78	72.14
DNN	71.30	71.20	71.67	72.03	71.04	71.33	71.48	71.80	70.96	71.12	71.46	71.51
SAGE	70.55	70.97	70.86	70.89	69.82	69.69	70.46	71.03	69.11	69.66	71.25	71.06
IGC	62.23	61.78	62.20	62.25	61.87	62.30	63.11	63.17	62.60	62.91	63.11	63.15
IGC+	73.15	73.37	73.92	74.34	73.87	73.99	74.22	74.51	74.14	74.22	74.53	74.79
MEIRec	64.94	65.10	65.30	65.53	65.45	65.55	65.66	65.72	65.19	65.58	66.20	65.63
MEIRec+	76.82	77.40	77.06	78.29	76.97	77.75	76.87	76.36	76.58	77.29	76.63	77.66
HGSRec$_{\backslash att}$	86.63	86.95	87.16	87.26	87.00	87.27	87.31	87.51	87.11	87.23	87.34	87.59
HGSRec$_{\backslash tra}$	78.17	79.10	79.50	79.95	76.40	79.12	77.09	79.63	78.22	78.89	78.83	81.37
HGSRec	**86.84**	**87.20**	**87.36**	**87.45**	**87.05**	**87.39**	**87.43**	**87.69**	**87.27**	**87.53**	**87.72**	**87.92**
Impro(%).	13.0	12.7	13.4	11.7	13.1	12.4	13.7	14.8	14.0	13.2	14.5	13.2

Table 2: The AUC comparisons of different methods.

based recommend model. Similar to IGC+, we also extend MEIRec as MEIRec+ for share recommendation.

• **HGSRec**$_{\backslash att}$: It is a variant of HGSRec, which removes the dual co-attention mechanism and employs the simple average strategy on all meta-paths for recommendation.

• **HGSRec**$_{\backslash tra}$: It is a variant of HGSRec, which removes the transitive triplet representation and concatenates the embeddings of u, i, v for recommendation.

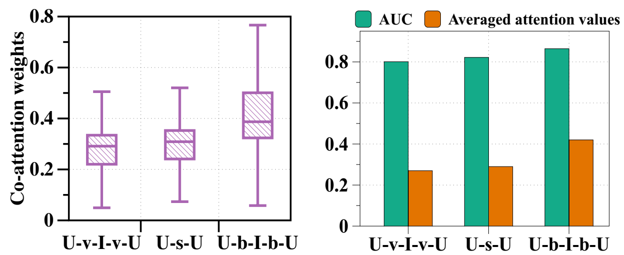
We select AUC as the evaluation metric, RMSProp as optimizer. We uniformly set feature embedding to 8, node embedding to 128, batch size to 1024, learning rate to 0.01 and dropout rate to 0.6 for deep models. For XGBoost, we set tree depth to 6, tree number to 10. For LR, we set the L1 reg to 1. For HeteGNNs, we sample 5, 10, 2 neighbors via *U-s-U*, *U-v-I-v-U* , *U-b-I-b-U* to learn multiple user embeddings and sample 50 neighbors via *U-b-I* to learn item embedding.

Performance Evaluation As shown in Table 2, we have the following observations: (1) HGSRec consistently performs better than all baselines with significant improvements. Compared to the best baseline, the improvements are up to 11.7%-14.5%, indicating the superiority of HGSRec. (2) Most of GNNs (i.e., GraphSAGE, IGC, and MEIRec) outperform feature based methods (i.e., LR, DNN, and XGBoost), indicating the importance of structure information. When deeper insight into these methods, we can find, if employing ternary interactions, the tripartite versions (i.e., IGC+ and MEIRec+) significantly outperform the original versions. It further confirms the benefits of modeling ternary interaction for share recommendation. (3) Comparing the performance of HGSRec with its variants, we can find HGSRec achieves the best performance. The degradation of HGSRec$_{\backslash att}$ indicates the effectiveness of the dual co-attention mechanism, while the degradation of HGSRec$_{\backslash tra}$ validates the superiority of transitive triplet representation. Note that the degradation of HGSRec$_{\backslash tra}$ is much more significant than that of HGSRec$_{\backslash att}$, which implies that transitive triple representation may make higher contribution than dual co-attention mechanism.

Attention Analysis The dual co-attention mechanism can dynamically fuse multiple embeddings of $User$ and $Friend$ with regard to different $Items$ and improve the share suitabilities. We first present the macro-level analysis via the box-plot figure of attention distributions over $User$ on 3-day dataset in Figure 4(a). Note that attention values distributions over $Friend$ also show similar phenomenons. As can be seen, the attention distribution of meta-paths are different, and the attention values of *U-b-I-b-U* is the largest with a higher variance, which illustrates that this meta-path is the most important for most users. The reason is that *U-b-I-b-U* is related to user purchasing behavior which reflects the strongest user preference. The higher variance of *U-b-I-b-U* also implies its importances varies greatly for different users. We further test HGSRec with single meta-path and show their performances with the corresponding averaged attention values in Figure 4(b). Consistent with attention distribution, *U-b-I-b-U* is the most useful meta-path which achieves the highest AUC and gets the largest attention value.

We further present a case study to show the potential interpretability of HGSRec. We select a share action $\langle u707, i586, v198\rangle$, where a user $u707$ shares an eye shadow $i586$ to his friend $v198$. Note that eye shadow $i586$ belonging to the category of $Makeup/Perfume$. As can be seen in Figure 5, the learned attention values between $\langle u707, i586\rangle$ and $\langle v198, i586\rangle$ are significantly different from each other. The interaction between $\langle u707, i586\rangle$ mainly depends on *U-s-U*, while the interaction between $\langle v198, i586\rangle$ mainly depends on *U-s-U* and *U-v-I-v-U*. By inspecting into the dataset, we found that most of $u707$'s friends like to buy $Makeup/Perfume$ and some of them buy more than 20 items belonging to this category. It explains why the *U-s-U* plays the key role in $\langle u707, i586\rangle$. The users who connect to $v198$ via *U-s-U* and *U-v-I-v-U* also like buying the items belonging to $Makeup/Perfume$ which explains why *U-s-U* and *U-v-I-v-U* both play the key roles in $\langle v198, i586\rangle$. In summary, the proposed HGSRec is able to learn appropriate attention values for user and friend with regard to different items in different share actions, providing potential interpretability for recommendation results.

Effects of Different Meta-paths To further investigate



(a) Attention distributions. (b) Performance & averaged attention values.

Figure 4: The attention analysis on 3-days dataset.

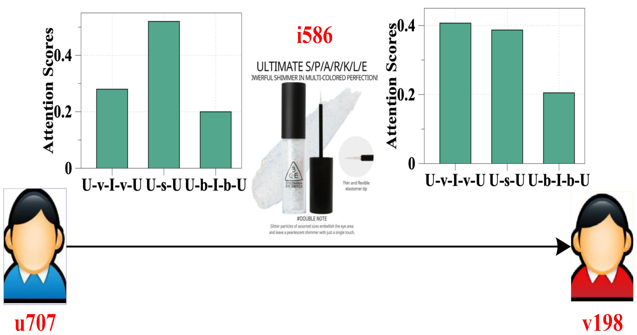


Figure 5: Attention values of meta-paths in a share action $\langle u707, i586, v198\rangle$ on 3-days dataset.

the effect of different meta-paths, we test the performance of HGSRec on 3-day dataset via adding four meta-paths (*U-b-I*, *U-v-I-v-U*, *U-b-I-b-U*, and *U-s-U*) one by one. As shown in Figure 6, with the addition of meta-paths, the performance of HGSRec improves consistently. It demonstrates more comprehensive information extracted by meta-paths indeed improves node embedding. Note that different meta-paths have different impacts. When adding *U-b-I-b-U*, HGSRec achieves the largest improvement. Similar to the results in Figure 4(a), *U-b-I-b-U* is the most informative path, because it directly reflects user intention.

Online Experiments We deploy HGSRec on Taobao APP for online share recommendation and compare HGSRec with XGBoost via online A/B testing. Online service need to satisfy the following requirements: (1) Storage and processing for massive data. Share recommendation system is stored on MaxCompute as adjacency list for memory efficiency. (2) Abnormal share action. We filter abnormal share actions (e.g., a user shares more than thousands of items with his friend within 24 hours). (3) New feature and missing feature. New features comes everyday, so we leverage hash function to map all features, leading a slight loss of performance when hash collision happens. Missing features are padded with a specific $token$.

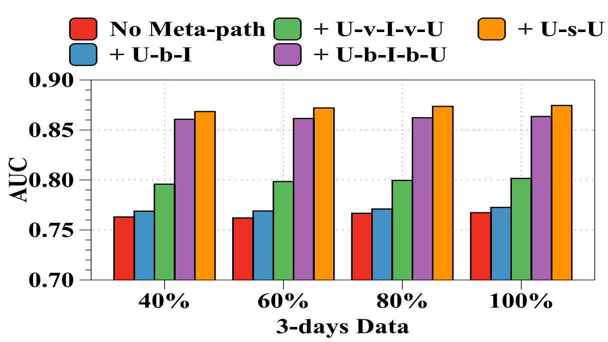


Figure 6: Performances of HGSRec (training ratio from 40% to 100%) with different numbers of meta-paths.

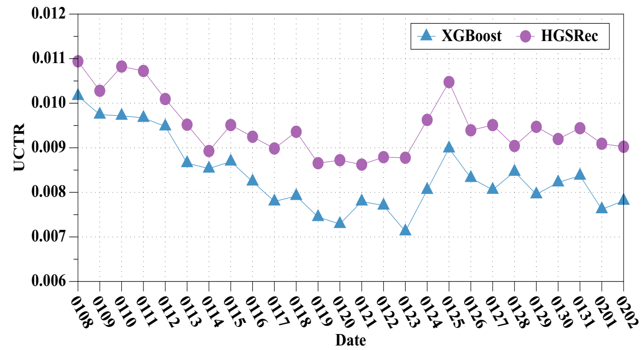


Figure 7: The results of Online A/B testing.

The online results range from 2020/01/08 to 2020/02/02 (25 days) are shown in Figure 7. Here we select UCTR (UCTR=Unique Click/Unique Visitor) for online evaluation. The larger UCTR, the better performance. The long-term observations show that HGSRec consistently outperforms XGBoost with a significant gap, demonstrating the high industrial practicability and stability of HGSRec.

Conclusion

In this paper, we first study the problem of share recommendation in social e-commerce, whose goal is to predict whether a user will share an item to his friend. We first construct an attributed heterogeneous graph to represent share scenario and propose a novel heterogeneous GNN based share recommendation model, called HGSRec. With the help of feature embedding and semantic aggregation, the proposed HGSRec learns multiple embeddings of u, i, v under different meta-paths via tripartite heterogeneous GNNs, and then dynamically fuses them via dual co-attention mechanism, followed by a transitive triplet representation to model the asymmetric share action. Extensive experiments on Taobao demonstrate the superiority of the proposed HGSRec.

Acknowledgements

This work is supported in part by the National Natural Science Foundation of China (No. U1936220, 61772082, 61702296, 62002029, U1936104, 61972442), and the BUPT Excellent Ph.D. Students Foundation (No. CX2020311).

References

Antoine, B.; Nicolas, U.; Alberto, G.-D.; Jason, W.; and Oksana, Y. 2013. Translating Embeddings for Modeling Multi-relational Data. In *NIPS*, 2787–2795.

Fan, S.; Zhu, J.; Han, X.; Shi, C.; Hu, L.; Ma, B.; and Li, Y. 2019a. Metapath-guided Heterogeneous Graph Neural Network for Intent Recommendation. In *KDD*, 2478–2486.

Fan, W.; Ma, Y.; Li, Q.; He, Y.; Zhao, E.; Tang, J.; and Yin, D. 2019b. Graph neural networks for social recommendation. In *WWW*, 417–426.

Gefen, D.; and Straub, D. W. 2004. Consumer trust in B2C e-Commerce and the importance of social presence: experiments in e-Products and e-Services. *Omega* 32(6): 407–424.

Hamilton, W. L.; Ying, R.; and Leskovec, J. 2017. Inductive Representation Learning on Large Graphs. In *NIPS*, 1024–1034.

Hu, B.; Shi, C.; Zhao, W. X.; and Yu, P. S. 2018. Leveraging Meta-path based Context for Top-N Recommendation with A Neural Co-Attention Model. In *KDD*.

Hu, L.; Yang, T.; Shi, C.; Ji, H.; and Li, X. 2019. Heterogeneous Graph Attention Networks for Semi-supervised Short Text Classification. In *EMNLP*, 4823–4832.

Kipf, T. N.; and Welling, M. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *ICLR*.

Kutty, S.; Nayak, R.; and Chen, L. 2014. A people-to-people matching system using graph mining techniques. In *WWW*, 311–349.

Ma, H.; Yang, H.; Lyu, M. R.; and King, I. 2008. Sorec: social recommendation using probabilistic matrix factorization. In *CIKM*, 931–941.

Ricci, F.; Rokach, L.; and Shapira, B. 2011. Introduction to recommender systems handbook. In *Recommender systems handbook*, 1–35. Springer.

Sarwar, B.; Karypis, G.; Konstan, J.; and Riedl, J. 2001. Item-based collaborative filtering recommendation algorithms. In *WWW*.

Shi, C.; Hu, B.; Zhao, X.; and Yu, P. 2018. Heterogeneous Information Network Embedding for Recommendation. *IEEE Transactions on Knowledge and Data Engineering* .

Sun, Y.; Han, J.; Yan, X.; Yu, P. S.; and Wu, T. 2011. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. In *VLDB*, 992–1003.

Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; and Bengio, Y. 2018. Graph Attention Networks. In *ICLR*.

Wang, X.; Bo, D.; Shi, C.; Fan, S.; Ye, Y.; and Yu, P. S. 2020. A Survey on Heterogeneous Graph Embedding: Methods, Techniques, Applications and Sources. *ArXiv* abs/2011.14867.

Wang, X.; He, X.; Wang, M.; Feng, F.; and Chua, T.-S. 2019a. Neural graph collaborative filtering. In *SIGIR*, 165–174.

Wang, X.; Ji, H.; Shi, C.; Wang, B.; Ye, Y.; Cui, P.; and Yu, P. S. 2019b. Heterogeneous Graph Attention Network. In *WWW*, 2022–2032.

Wang, Z.; Liao, J.; Cao, Q.; Qi, H.; and Wang, Z. 2014. Friendbook: a semantic-based friend recommendation system for social networks. *IEEE transactions on mobile computing* 14(3): 538–551.

Wu, L.; Sun, P.; Fu, Y.; Hong, R.; Wang, X.; and Wang, M. 2019. A neural influence diffusion model for social recommendation. In *SIGIR*, 235–244.

Ying, R.; He, R.; Chen, K.; Eksombatchai, P.; Hamilton, W. L.; and Leskovec, J. 2018. Graph convolutional neural networks for web-scale recommender systems. In *KDD*, 974–983.

Zhao, J.; Zhou, Z.; Guan, Z.; Zhao, W.; Ning, W.; Qiu, G.; and He, X. 2019. IntentGC: a Scalable Graph Convolution Framework Fusing Heterogeneous Information for Recommendation. In *KDD*, 2347–2357.