Learning Consumer and Producer Embeddings for User-Generated Content Recommendation

Wang-Cheng Kang
UC San Diego
wckang@ucsd.edu

Julian McAuley
UC San Diego
jmcauley@ucsd.edu

ABSTRACT

User-Generated Content (UGC) is at the core of web applications where users can both produce and consume content. This differs from traditional e-Commerce domains where content producers and consumers are usually from two separate groups. In this work, we propose a method *CPRec* (consumer and producer based recommendation), for recommending content on UGC-based platforms. Specifically, we learn a core embedding for each user and two transformation matrices to project the user's core embedding into two 'role' embeddings (i.e., a producer and consumer role). We model each interaction by the ternary relation between the consumer, the consumed item, and its producer. Empirical studies on two large-scale UGC applications show that our method outperforms standard collaborative filtering methods as well as recent methods that model producer information via item features.

KEYWORDS

Collaborative Filtering, User-Generated Content Recommendation

ACM Reference Format:
Wang-Cheng Kang and Julian McAuley. 2018. Learning Consumer and Producer Embeddings for User-Generated Content Recommendation. In *Twelfth ACM Conference on Recommender Systems (RecSys '18), October 2–7, 2018, Vancouver, BC, Canada.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3240323.3240394

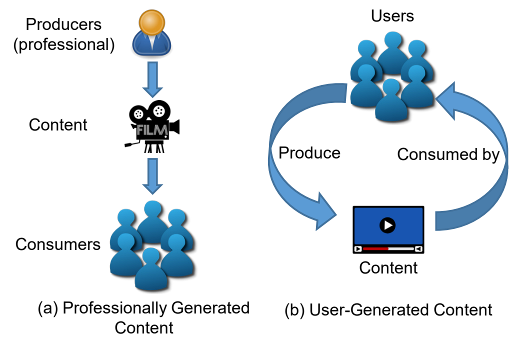


Figure 1: An illustration of the production and consumption processes of professionally generated content versus UGC. In centralized domains, producers and consumers form two separate groups. In contrast, users simultaneously perform the two roles (i.e., being a 'prosumer') in UGC platforms.

1 INTRODUCTION

We consider the problem of providing recommendations on user-generated content (UGC) communities. Unlike traditional domains for recommendation, UGC communities form a 'closed loop' between users and items (Figure 1). That is, each user can have two roles: a consumer and a producer. However, conventional recommender systems in centralized domains (e.g. Amazon, Netflix) only focus on users' consumption behavior (e.g. clicks, purchases, views, etc.), i.e., the relationship between the consumer and the item. However for UGC applications we have additional dynamics to model, namely the ternary relationship between the consumer, the item, and the item's producer (who is in turn also a user).

Our main goal in this paper is to design models for recommendation that explicitly consider this ternary dynamic. Specifically, we propose a method *CPRec* (consumer and producer based recommendation), which learns two role embeddings derived from the same core embedding via two projection matrices. We model users' consumption behavior by the summation of her preference toward the item and her 'appreciation' toward the item's producer. We compare our method against various baselines on two UGC platforms (*Pinterest* and *Reddit*).

RecSys '18, October 2–7, 2018, Vancouver, BC, Canada

ACM ISBN 978-1-4503-5901-6/18/10.
https://doi.org/10.1145/3240323.3240394

2 RELATED WORK

Recommender Systems: Recommender systems focus on modeling the compatibility between users and items, based on historical feedback (e.g. clicks, purchases, likes). Matrix Factorization (MF) methods seek to uncover latent dimensions to represent users' preferences and items' properties, and estimate user-item interactions through the inner product between the user embedding and item embedding [6, 13]. User feedback can be *explicit* (e.g. ratings) or *implicit* (e.g. clicks, purchases, comments) [3, 12]. Modeling implicit feedback can be challenging due to the ambiguity of interpreting 'non-observed' (e.g. non-purchased) data. Recently, *point-wise* [3] and *pairwise* [12] methods are proposed to solve such challenges.

Point-wise methods assume non-observed feedback to be inherently negative, and model the problem via regression, either by assigning 'confidence levels' to feedback [3], or by sampling non-observed feedback as negative instances [9].

Pairwise methods are based on a weaker but possibly more realistic assumption that positive feedback must only be 'more preferable' than non-observed feedback. Such methods directly optimize the ranking (in terms of the AUC) of the feedback and are to our knowledge state-of-the-art for implicit feedback datasets. In particular, Bayesian Personalized Ranking (BPR), has experimentally been shown to outperform a variety of competitive baselines [12].

In this work, we treat users' consuming behaviors as implicit feedback, and seek to optimize their personalized pairwise ranking.

Ownership-Aware Recommendation: Other than standard MF approaches for estimating user-item interactions, Factorization Machines (FMs) [11] provide a generic factorization approach that can easily incorporate side information of users and items. In our case, producer information can be viewed as an item feature (via a one-hot representation), that can be used by FMs. Recently, Vista [2] was proposed for artistic recommendation with ownership information. Though the two methods can make use of ownership information, they do not specifically model the two roles (consumer and producer) of each user as we do. We discuss these methods in more detail in the next section, and compare our method against them empirically.

Socially-Aware Recommendation: Leveraging social networks can help us understand user-user relationships and improve the performance of item recommendation [7, 8, 10, 16]. The social network is usually based on friendship or 'trust' relationships. A typical assumption in socially-aware methods is that users prefer to follow their friends' behaviors since they may share common interests. However, in our problem setting, there is not such an explicit social network between users, though our model tries to uncover a implicit 'follow' relationships from users' consumption and production patterns.

Heterogeneous User Feedback Modeling: Unlike conventional recommendation methods that only consider a single type of feedback, there is a line of work that seeks to model users' different types of behavior. For example, DualRec [14] considers a reviewer role and a rater role (rates the helpfulness of reviews). A recent method SPTF [15] is a more general approach to jointly model different types of user behavior (e.g. click, add-to-cart, purchase) via probabilistic tensor factorization. However, these methods typically focus on modeling users' consumption behaviors, whereas we consider both consumer and producer behaviors in UGC platforms.

3 CPREC: CONSUMER AND PRODUCER BASED RECOMMENDATION

3.1 Problem Description

We consider a system that makes recommendations on UGC applications with implicit feedback (e.g. click, comment, retweet, etc.). Since there is no observed negative feedback, the goal is to rank items such that 'observed' items should be ranked higher than non-observed items. A critical property in this domain is that users not only provide feedback on items, but also that all items are created by the users themselves. That is, each user assumes both the role of a consumer and producer. We use $\mathcal{U}$ and $\mathcal{I}$ to represent the set of users and items (respectively). For each user u, we use $\mathcal{I}_u^+$ to denote all items toward which she has provided positive feedback. Finally, each item $i \in \mathcal{I}$ is produced by the user $p_i \in \mathcal{U}$.

We define the sets $C \subset \mathcal{U}$ and $P \subset \mathcal{U}$ to represent consumers (who provided any feedback) and producers (who created any item). In addition to $C \cap P = \mathcal{U}$, we have $PS = C \cup P$ for representing 'prosumers' who both created and consumed items. The ratio $|PS|/|\mathcal{U}|$ is critical to identify how the groups of consumers and producers overlap. Table 1 summarizes our notation.

Table 1: Notation.

Notation	Explanation
$\mathcal{U}, \mathcal{I}$	user and item set
$\mathcal{I}_u^+$	positive item set for user u
$p_i \in \mathcal{U}$	the producer of item i
$\widehat{x}_{ui} \in \mathbb{R}$	predicted score user u gives to item i
$K \in \mathbb{N}$	latent factor dimensionality
$\boldsymbol{\gamma}_u \in \mathbb{R}^K$	core embedding for user u
$\mathbf{W}_u^c, \mathbf{W}_u^p \in \mathbb{R}^{K\times K}$	role transformation matrices (consumer and producer)
$\boldsymbol{\gamma}_i \in \mathbb{R}^K$	item i's embedding
$C \subset \mathcal{U}$	set of all consumers
$P \subset \mathcal{U}$	set of all producers
$PS = C \cup P$	set of all prosumers

3.2 The CPRec Model

Biased matrix factorization is widely used as an underlying preference predictor in recommendation problems [5, 12]. Specifically, it models user-item interactions via bias terms and an inner product between the latent vectors of the user and item:

$$\widehat{x}_{ui} = \alpha + \beta_u + \beta_i + \langle \boldsymbol{\gamma}_u, \boldsymbol{\gamma}_i \rangle$$

Though this has shown strong performance in modeling user-item interactions, it does not fully model ternary interactions between the consumer u, the item i, and its producer p_i. Hence we propose a model to capture this interaction by factorization into two parts:

$$\widehat{x}_{ui} = \alpha + \beta_u + \beta_i + \underbrace{\langle \boldsymbol{\gamma}_u^c, \boldsymbol{\gamma}_i \rangle}_{\text{consumer-item preference}} + \overbrace{\langle \boldsymbol{\gamma}_u^c, \boldsymbol{\gamma}_{p_i}^p \rangle}^{\text{consumer-producer appreciation}}.$$

We introduce two embeddings ($\boldsymbol{\gamma}_u^c$ and $\boldsymbol{\gamma}_u^p$) to represent the user u's two roles (consumer and producer). This is mainly because: (1) The 'follow' relationship between users is asymmetric, which cannot be modeled by the inner product of homogeneous embeddings (i.e., using the same user embedding for her two roles); 2) Users may exhibit different behavior when they play different roles. Ultimately, we model user u's consumer embedding and producer embedding as being derived from a single core embedding $\boldsymbol{\gamma}_u$ and two transformation matrices:

$$\boldsymbol{\gamma}_u^c = \mathbf{W}^c \boldsymbol{\gamma}_u, \qquad \boldsymbol{\gamma}_u^p = \mathbf{W}^p \boldsymbol{\gamma}_u$$

where $\mathbf{W}^c, \mathbf{W}^p \in \mathbb{R}^{K\times K}$. That is to say, we use two projection matrices to project a user's core embedding into her two role embeddings. The advantage here is we only introduce $2K^2$ new parameters (compared to standard MF) to achieve asymmetric embeddings, which is helpful to avoid overfitting.

3.3 Learning

Based on the proposed preference predictor, we adopt a Bayesian Personalized Ranking (BPR) framework [12] to learn all parameters.

The goal of BPR is to approximately optimize the AUC of ranking observed feedback for each user. Specifically, we consider the triplets $(u,i,j) \in \mathcal{D}$, where:

$$\mathcal{D} = \{(u,i,j) | u \in \mathcal{U} \wedge i \in \mathcal{I}_u^+ \wedge j \in \mathcal{I} \setminus \mathcal{I}_u^+\}.$$

Here $i \in \mathcal{I}_u^+$ is an item about which the user u has provided feedback, whereas $j \in \mathcal{I} \setminus \mathcal{I}_u^+$ is one about which they have not. Thus intuitively, for a user u, the predictor should assign a larger preference score to item i than item j. Hence BPR defines the difference between preference scores by

$$\widehat{x}_{uij} = \widehat{x}_{ui} - \widehat{x}_{uj}.$$

Note that the global bias term α and user bias term β_u are naturally canceled in $\widehat{x}_{uij}$. We seek to optimize the ranking by maximizing the posterior

$$\begin{aligned}\ln p(\Theta|\mathcal{D}) &\propto \ln \prod_{(u,i,j)\in\mathcal{D}} \sigma(\widehat{x}_{uij}) p(\Theta)\\ &= \sum_{(u,i,j)\in\mathcal{D}} \ln \sigma(\widehat{x}_{uij}) - \lambda_\Theta \|\Theta\|^2,\end{aligned}$$

where $\sigma(\cdot)$ is the sigmoid function, $\Theta = \{\boldsymbol{\gamma}_u, \boldsymbol{\gamma}_i, \beta_i, \mathbf{W}^c, \mathbf{W}^p\}$ includes all model parameters, and λ_Θ is a regularization hyperparameter. We adopt the Adam optimizer [4], a variant of stochastic gradient descent with adaptive estimation of moments, to learn all variables.

3.4 Discussion of Related Methods

Vista [2] is a recent method for artistic recommendation (applied to data from *behance.net*), which leverages ownership information.[1] Vista models user-item interactions via:

$$\widehat{x}_{ui} = \langle \boldsymbol{\gamma}_u^{(1)}, \boldsymbol{\gamma}_i \rangle + \langle \boldsymbol{\gamma}_u^{(2)}, \boldsymbol{\gamma}_{p_i}^{(2)} \rangle.$$

One major difference is that Vista uses the inner product of symmetric embeddings to model relationships between users. However, as we stated before, the "follow" relationship between users may be asymmetric in the domains we consider. Hence our model may be more suitable for capturing the relationship between users.

Another way to use producer information is to view it as a categorical item feature. Factorization Machines (FMs) provide a generic factorization method by modeling interactions between users, items, and their features. Here, we assign each item i a one-hot feature $\mathbf{f}_i \in \{0,1\}^{|\mathcal{U}|}$, and $\mathbf{f}_{p_i} = 1$. The (second-order) estimator of FMs is given by:

$$\widehat{x}_{ui} = \beta_u + \beta_i + \beta_{p_i} + \langle \boldsymbol{\gamma}_u^{(1)}, \boldsymbol{\gamma}_i \rangle + \langle \boldsymbol{\gamma}_u^{(1)}, \boldsymbol{\gamma}_{p_i}^{(2)} \rangle + \langle \boldsymbol{\gamma}_i, \boldsymbol{\gamma}_{p_i}^{(2)} \rangle.$$

FMs can capture asymmetric user relationships since they make use of two different embeddings (i.e., $\boldsymbol{\gamma}_u^{(1)}, \boldsymbol{\gamma}_{p_i}^{(2)}$) to model their interaction. However, it is not aware that $\boldsymbol{\gamma}_u^{(1)}$ and $\boldsymbol{\gamma}_u^{(2)}$ are the same individual. We argue that the two embeddings should be connected since users' behaviors under the two roles should be related (yet different).

Other than the above, there are a few methods that focus on 'socially-aware' recommendation, such as Social-BPR [16]. However, our problem setting differs from that of social recommendation as we don't have an explicit social network of users, but rather we focus on modeling interactions which are related to (u,i,p_i) where p_i is the creator of item i.

[1] Vista also considers visual and temporal information, however they are not the focus of this paper.

Table 2: Dataset statistics (after preprocessing)

	Pinterest	Reddit
#users ($|\mathcal{U}|$)	134,747	52,654
#items ($|\mathcal{I}|$)	201,792	336,743
#actions ($\sum_{u\in\mathcal{U}} |\mathcal{I}_u^+|$)	690,506	1,786,032
consumer ratio ($|C|/|\mathcal{U}|$)	93.65%	99.60%
producer ratio ($|P|/|\mathcal{U}|$)	80.76%	87.24%
prosumer ratio ($|PS|/|\mathcal{U}|$)	74.42%	86.85%

4 EMPIRICAL STUDIES

4.1 Dataset

We consider two public datasets from UGC-based applications:

- **Pinterest** is a content discovery application based mainly around images. People can browse, upload (pin), like, and save (repin) images. We use the dataset crawled by [17],[2] which includes 0.89M users, 2.4M images and 56M actions ('save' and 'like') in January 2013 . We treat both "like" and "repin" actions as implicit feedback. Each item is associated with an uploader.
- **Reddit** is a discussion website which covers a variety of topics including news, science, movies, etc. Users can submit content and comment on submissions. We use a dataset which includes all submissions and comments on Reddit in March 2017.[3] Specifically, the dataset includes 1.3M users, 9.6M submissions 48M comments. We view each submission as an item, and commenting actions as implicit feedback. Each submission is associated with a single author.

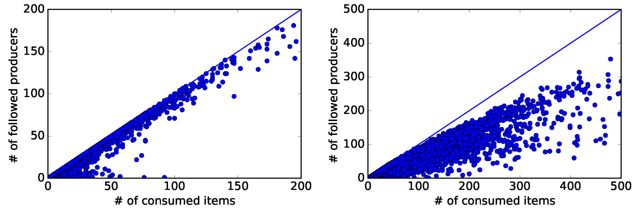
For data preprocessing, we discard inactive users and items which have fewer than 10 associated actions. For each user, we randomly withhold one action for validation $\mathcal{V}_u$, and another for testing $\mathcal{T}_u$. All remaining items are used for training $\mathcal{P}_u$, and we always tune models via our validation set and report the performance on the test set. Table 2 lists statistics of our datasets. Figure 2 shows users repeatedly following appreciated producers. We also considered the *behance* dataset used in [2], however, it proved unsuitable for our problem since its prosumer ratio is almost zero, which indicates that most of its users only have a single role (consumer or producer).

4.2 Baselines

We compare our method with standard recommendation methods as well as methods that make use of producer information.

- **PopRec** is a straightforward baseline which ranks items according to their popularity.
- **Bayesian Personalized Ranking (BPR)** [12] is a state-of-the-art recommendation method for implicit feedback. We use biased matrix factorization as the underlying predictor ($\widehat{x}_{ui} = \beta_i + \langle \boldsymbol{\gamma}_u, \boldsymbol{\gamma}_i \rangle$).

[2] https://nms.kcl.ac.uk/nishanth.sastry/projects/cd-gain/dataset.html
[3] https://redd.it/6607j2



(a) Pinterest (b) Reddit

Figure 2: Number of followed producers vs. number of consumed items. Each point represents a user. If user u consumed N items that are created from N different producers, her point in the figure will lie on the line. This shows that Reddit users are more inclined to repeatedly follow the same producer.

- **Factorization Machines (FMs)** [11] provide a generic factorization approach that can be used to model interactions between users, items, and their features. We use a one-hot encoding to represent the producer of each item.
- **Visually, Socially, and Temporally-Aware Recommendation (Vista)** [2] is a recent method for artistic recommendation. We use a reduced model which only considers ownership information.

These baselines are intended to show (a) the importance of learning personalized notions of compatibility (MF methods vs. PopRec); (b) the effect of modeling consumer-producer interactions for UGC applications (FM/Vista/CPRec vs. BPR); and (c) the improvement gained by our consumer/producer embedding approach (CPRec vs. FM/Vista).

We implemented all MF methods using *Tensorflow* [1]. For fair comparison, we train all methods using the BPR loss and Adam optimizer [4]. We tune hyperparameters via grid search on a validation set, with a regularizer selected from {0.001, 0.01, 0.1, 1}. The learning rate is set to 0.01, and the batch size is 10000. The code is available at https://github.com/kang205/CPRec.

4.3 Recommendation Performance

We measure the recommendation performance via the AUC (Area Under the ROC Curve) [12], which considers the overall ranking:

$$AUC = \frac{1}{|\mathcal{U}|}\sum_{u\in\mathcal{U}}\frac{1}{|\mathcal{D}_u|}\sum_{(i,j)\in\mathcal{D}_u}\xi(\widehat{x}_{ui} > \widehat{x}_{uj}),$$

where $\mathcal{D}_u = \{(i,j)|(u,i) \in \mathcal{T}_u \wedge (u,j) \notin (\mathcal{P}_u \cup \mathcal{V}_u \cup \mathcal{T}_u)\}$ and $\xi(\cdot)$ is an indicator function. Intuitively, AUC is the fraction of times that the 'observed' items i are ranked higher than 'non-observed' items j.

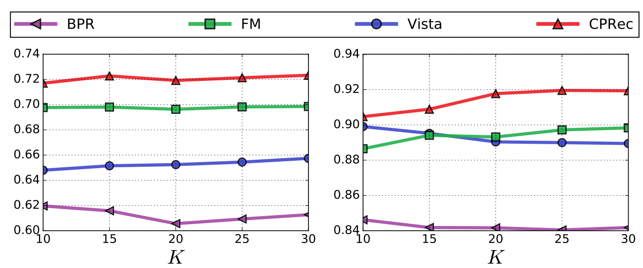
We consider two recommendation target groups: all consumers and cold consumers (who have fewer than five consumed items). We do not consider content recommendation for users who only produced items since we lack ground truth items for evaluation. Table 3 shows the ranking performance of all methods on the two datasets with latent dimensionality $K = 20$. The overall performance on Reddit is better than that of Pinterest, a possible reason being that Reddit users are more inclined to repeatedly consume content by the same producers (as shown in Figure 2) which makes their behaviors easier to predict.

We can see that FM/Vista/CPRec are more accurate than BPR (especially for cold users), which shows the importance of considering producer information in UGC applications. Moreover, *CPRec* outperforms all other baselines in all the settings. Compared to BPR, *CPRec* gains 13.9% AUC improvement for all users and 18.6% AUC improvement for cold users on average. Our method also achieves a 2.8% improvement against the strongest baseline on average.

To examine the effect of the latent dimensionality K which directly relates to model complexity, we plot the ranking performance for increasing K for all MF methods in Figure 3. We can see that our method *CPRec* is consistently better than the baselines with different K. Especially on Reddit, as K increases, the performance gap between *CPRec* and the strongest baseline becomes wider. However, we find that *CPRec* can achieve satisfactory performance with $K = 20$ on both datasets.

Table 3: Recommendation performance in terms of the AUC with latent dimensionality $K = 20$.

Dataset	Target	(a) PopRec	(b) BPR	(c) FM	(d) Vista	(e) CPRec
Pinterest	All users	0.6125	0.6056	0.6963	0.6524	0.7191
	Cold users	0.5993	0.5704	0.6855	0.6316	0.6902
Reddit	All users	0.6397	0.8416	0.8931	0.8903	0.9177
	Cold users	0.5564	0.7727	0.8668	0.8535	0.8980



(a) Pinterest (b) Reddit

Figure 3: Effect of the latent dimensionality K. Ranking performance (AUC) of all consumers is shown.

5 CONCLUSIONS AND FUTURE WORK

In this work, we consider recommendation on user-generated content (UGC) communities, and design a recommendation method *CPRec*, which learns two role embeddings (for consumer and producer roles) derived from the same core user embedding. We model each interaction via a ternary relation between the consumer, the consumed item, and its producer. We analyze the difference between *CPRec* and related methods. Empirically, extensive results on two UGC platforms demonstrate the effectiveness of our method. In the future, we plan to further investigate the problem of incorporating more context information and side features, including explicit social networks and temporal dynamics.

REFERENCES

[1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. 2016. TensorFlow: A System for Large-Scale Machine Learning.. In *OSDI'16*.

[2] Ruining He, Chen Fang, Zhaowen Wang, and Julian McAuley. 2016. Vista: A visually, socially, and temporally-aware model for artistic recommendation. In *RecSys'16*.

[3] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative filtering for implicit feedback datasets. In *ICDM'08*.

[4] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *ICLR'15*.

[5] Yehuda Koren. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *SIGKDD'08*.

[6] Yehuda Koren and Robert Bell. 2011. Advances in Collaborative Filtering. In *Recommender Systems Handbook*. Springer.

[7] Artus Krohn-Grimberghe, Lucas Drumond, Christoph Freudenthaler, and Lars Schmidt-Thieme. 2012. Multi-relational matrix factorization using bayesian personalized ranking for social network data. In *WSDM'12*.

[8] Hao Ma, Dengyong Zhou, Chao Liu, Michael R. Lyu, and Irwin King. 2011. Recommender systems with social regularization. In *WSDM'11*.

[9] Rong Pan, Yunhong Zhou, Bin Cao, Nathan N Liu, Rajan Lukose, Martin Scholz, and Qiang Yang. 2008. One-class collaborative filtering. In *ICDM'08*.

[10] Dimitrios Rafailidis and Fabio Crestani. 2017. Learning to Rank with Trust and Distrust in Recommender Systems. In *RecSys'17*.

[11] Steffen Rendle. 2010. Factorization Machines. In *ICDM'10*.

[12] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *UAI'09*.

[13] Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul Kantor. 2011. *Recommender systems handbook*. Springer US.

[14] Suhang Wang, Jiliang Tang, and Huan Liu. 2015. Toward Dual Roles of Users in Recommender Systems. In *CIKM'15*.

[15] Hongzhi Yin, Hongxu Chen, Xiaoshuai Sun, Hao Wang, Yang Wang, and Quoc Viet Hung Nguyen. 2017. SPTF: A Scalable Probabilistic Tensor Factorization Model for Semantic-Aware Behavior Prediction. In *ICDM'17*.

[16] Tong Zhao, Julian J. McAuley, and Irwin King. 2014. Leveraging Social Connections to Improve Personalized Ranking for Collaborative Filtering. In *CIKM'14*.

[17] Changtao Zhong, Sunil Shah, Karthik Sundaravadivelan, and Nishanth Sastry. 2013. Sharing the Loves: Understanding the How and Why of Online Content Curation. In *ICWSM'13*.