Mining Cross-network Association for YouTube Video Promotion

Ming Yan[1,2], Jitao Sang[1,2], Changsheng Xu[1,2]
[1]National Lab of Pattern Recognition, Institute of Automation, CAS, Beijing 100190, China
[2]China-Singapore Institute of Digital Media, Singapore, 139951, Singapore
{ming.yan, jtsang, csxu}@nlpr.ia.ac.cn

ABSTRACT

We introduce a novel cross-network collaborative problem in this work: given YouTube videos, to find optimal Twitter followees that can maximize the video promotion on Twitter. Since YouTube videos and Twitter followees distribute on heterogeneous spaces, we present a cross-network association-based solution framework. Three stages are addressed: (1) heterogeneous topic modeling, where YouTube videos and Twitter followees are modeled in topic level; (2) cross-network topic association, where the overlapped users are exploited to conduct cross-network topic distribution transfer; and (3) referrer identification, where the query YouTube video and candidate Twitter followees are matched in the same topic space. Different methods in each stage are designed and compared by qualitative as well as quantitative experiments. Based on the proposed framework, we also discuss the potential applications, extensions, and suggest some principles for future heterogeneous social media utilization and cross-network collaborative applications.

Categories and Subject Descriptors

H.3.5 [**Online Information Services**]: Web-based services

General Terms

Theory

Keywords

video promotion; cross-network analysis; social media

1. INTRODUCTION

Since the launch in 2005, YouTube has established itself as the world's largest video sharing platform. Latest statistics show that within every minute, 100 hours of video are uploaded to YouTube [1], resulting in an estimate of more than 2 billion videos totally. People act on purpose. It has been recognized that YouTube users share videos with an obvious extrinsic motivation of receiving attentions (e.g., video view) [1, 2], especially for the profit-seeking video content providers [2]. In spite of the fact that billions of videos are consumed in YouTube each day, the massive volume makes the exploration of individual videos very difficult. According to research, YouTube video view count distribution exhibits a power-law pattern with truncated tails [3]. Most videos have a short active life span, receiving half of the total views in the first 6 days after being published, and with fewer and fewer access thereafter [4]. Therefore, the mismatch between high attention expectation and rare access opportunity calls for YouTube video promotion to broaden the viewership.

Generally speaking, within YouTube, video can be accessed from internal search, related video recommendation, channel subscription or front page highlight. Some work has been devoted to utilizing these sources to promote internal video views. Zhou et al. studied the impact of related video recommendation on video views, with goal to design a strategy to drive YouTube video popularity [5]. In [6], YouTube search bias phenomenon is investigated to optimize video discovery in YouTube's internal search results. However, essentially as a content repository, YouTube exhibits limited promotion efficiency with the internal mechanisms. Very recent research shows that external referrers, such as external search engines and other social media websites, arise to be important sources to lead users to YouTube videos [7]. Among the social media websites, Twitter has been quickly growing as the top referrer source for web video discovery [3].

Twitter allows users to embed videos in their tweets by posting video links. Followers to these users then receive the tweet feed and become the potential viewers of these videos. The followee-follower architecture has established Twitter as a great platform to promote and engage with the audiences and distinguished itself with the significant information propagation efficiency. Twitter followees, especially those with a lot of followers (which we refer to as *popular followee*), play important roles under social media circumstances by: (1) acting as "we media", via the control of information dissemination channels to millions of audiences, and (2) acting as influential leaders, via their potential impact on the followers' decisions and activities. YouTube video "Gangnam Style" went viral to become the first web video

[1] http://www.youtube.com/yt/press/statistics.html.

[2] YouTube has started to let video content providers be partners to cash in on the videos posted by sharing ad revenue and charging rental fees to viewers.

[3] http://mashable.com/2010/05/25/twitter-online-video/.

MM'14, November 03 - 07, 2014, Orlando, FL, USA.

http://dx.doi.org/10.1145/2647868.2654920.

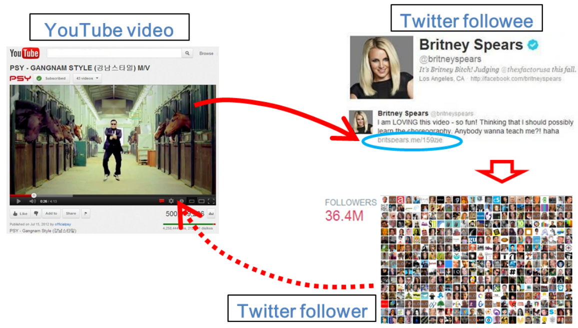

Figure 1: Problem illustration.

that reaches one billion views in 5 months, resulting mainly from its successful strategy of roping in some popularly followed musicians on Twitter, such as Britney Spears, Justin Bieber and Katy Perry. In this context, if we can identify "proper" followees to help disseminate videos, their significant audience accessibility and behavioral impact will guarantee the promotion efficiency. Therefore, the problem of this work is: *For specific YouTube video, to identify proper Twitter followees with goal to maximize video dissemination to the followers* (as shown in Fig. 1).

It is not trivial to measure the "properness" of Twitter followees for specific YouTube videos. The challenge lies in two-fold: (1) The level of "properness" is not necessarily proportional to the number of followers (#follower). While a popular followee with a large #follower will guarantee a huge audiences, what video promotion cares is the number of "effective" audiences, who are likely to show interest to the video and with higher probability to take subsequent consuming actions like watch, reshare, etc. A close analogy to advertising can be made, where the followee is viewed as advertising media, whose bid price is decided by #follower. Twitter followee identification is analogous to advertising media selection [4], with goal to achieve the maximum coverage and exposures in a target audience with the minimum cost. (2) Based on the above discussion, whether a Twitter followee is proper for the promotion task is actually decided by the interest his/her followers show to the YouTube videos. However, we only know the followers' activities on Twitter, based on what only the demographics or interests on the general level can be inferred [8, 9]. While, the YouTube videos are known to distribute more on specific semantic level [10]. The discrepancy in topic granularity and affiliated platform makes it impractical to directly evaluate Twitter followers' interest to YouTube videos, let alone the costly computation in evaluating each follower and the subsequent aggregation.

Our solution to address the above challenges is inspired by the fact that the same individual usually involves with different social media networks, including media sharing *YouTube* and *Flickr*, microblogging *Twitter* and *Tumblr*, private/professional social networks *LinkedIn* and *Facebook*. Anderson Analytics shows that the different social media networks share remarkable percentage of overlapped users [5]. In this context, if we know the corresponding Twitter accounts of YouTube users who show interest to a given video (e.g., upload, favorite, add to playlist), it is confident to identify the Twitter followee that these Twitter accounts jointly followed as the optimal promotion referrer. In practice, it is impossible to obtain all the overlapped accounts between different networks. Moreover, a practical solution should be not limited to the specific video and followee, but generalizable on the alike sets. Therefore, in this work, we propose to investigate the problem in YouTube video and Twitter followee topic level, and exploit the observed overlapped users to mine the cross-network topic association for solution. Specifically, based on users' interactions with YouTube videos and Twitter followees, we first build heterogeneous video topic and followee topic, respectively. After that, the topic association is mined from the overlapped users' distributions on the two topics. Finally, the optimal Twitter promotion referrers are identified by matching with the transferred video distribution on the Twitter followee topic space.

Our contributions in this work can be summarized in the following three-fold:

1. We introduce a new problem of YouTube video promotion on Twitter platform by identifying proper Twitter followees. There exist both trends and demands in exploring external referrers towards promoting social media content.
2. A cross-network association-based solution framework is presented, under which alternative methods have been examined. The solution is validated to discover heterogeneous topic association and facilitate effective video-followee matching in the same topic space.
3. The discussion in Section 5 on the idea of exploiting overlapped users' activities in different networks towards cross-network knowledge mining opens up possibilities to the utilization of heterogeneous social media sources. This will be the key takeaway for future cross-network analysis and applications.

2. RELATED WORK

2.1 Cross-network Collaboration

With various social media networks growing in prominence, netizens are using a multitude of social media services for social connection and information sharing. Cross-network collaborative applications have recently attracted attentions. One line is on cross-network user modeling, which focuses on integrating various social media activities. In [11], the authors introduced a cold-start recommendation problem by aggregating user profiles in Flickr, Twitter and Delicious. Deng et al. has proposed a personalized YouTube video recommendation solution by incorporating user information from Twitter [12]. Another line is devoted to taking advantage of different social networks' characteristics towards collaborative applications. For example, Suman et al. exploited the real-time and socialized characteristics of the Twitter tweets to facilitate video applications in YouTube [13]. Our work belongs to the second line, where a collaborative application is designed to exploit the propagation efficiency of Twitter to meet the YouTube video promotion demand.

2.2 Social Media Influencer Mining

Previous analysis on Twitter has found that popular users with high in-degree are not necessarily influencers for propagation [14], which calls for research onto the problem of

[4] http://en.wikipedia.org/wiki/Advertising media selection.
[5] See "Anderson Analytics 2009 report: what your favorite social network says about you?".

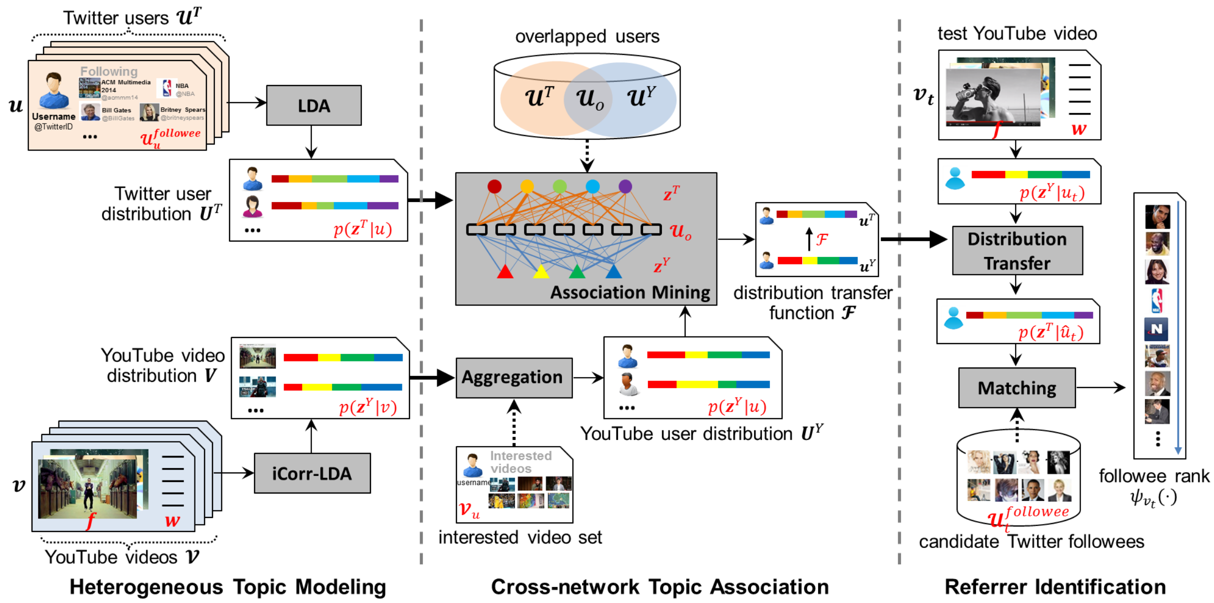

Figure 2: Solution framework.

influencer mining. One line is to identify the domain or topic experts. Representative solutions include the extensions to PageRank by considering topical similarity, e.g, TwitterRank [15], and incorporating auxiliary sources like Twitter lists [16]. Another line is concerned with maximizing influence spread by initializing some seed users. David et al. first defined this problem [17], which is then applied to viral marketing [18].

Our introduced problem of Twitter followee identification can be viewed as a special case of influencer mining. The existing influencer mining methods mainly focus on single network and need an explicit relevance metric, e.g. the topical relevance between follower and followee, and the accept rate between the propagation item and follower. In our problem, the relevance of influencer is designed by items distributed on another network. It is difficult to explicitly define the relevance metric between cross-network knowledge. Moreover, to focus on addressing cross-network association, we pay no attention to the complicated social network structure as in the standard maximizing influence problems. What we care is actually about the propagation efficiency in the first level of followee-follower network.

2.3 Heterogeneous Topic Association

The core of our solution lies in the heterogeneous topic association between Twitter followee and YouTube video. Typical applications of existing heterogeneous topic association work include cross-media retrieval and heterogeneous face recognition, where invariant feature extraction and subspace learning based solutions are extensively investigated. Invariant feature extraction methods are devoted to reducing the heterogeneous gap by exploring the most insensitive feature patterns. Klare et al. proposed to extract the SIFT and Multiscale LBP for forensic sketch and mug shot photo matching [19]. In [20], the intra-difference and inter-difference are jointly considered into a discriminant local feature learning framework. The basic idea of subspace learning is to learn a new space where the observed heterogeneous data can be well represented. [21] provides good surveys of CCA and its extensions to learn a semantic representation from multimodal data.

Subspace learning methods focus on maintaining the smoothness for retrieval, i.e., the projected coefficients of two items should be similar if they constitute a training pair. This is different from our goal for heterogeneous topic association and transfer. Invariant feature extraction aims to extract and learn low-level discriminative features, which will largely fail in case of complicated association like heterogeneous social media topics. In this work, we propose a solution framework based on user's collaborative involvement in heterogeneous topics. This avoids low-level analysis and can be viewed as a high-level crowdsourcing strategy.

3. CROSS-NETWORK YOUTUBE VIDEO PROMOTION

This section introduces the cross-network YouTube video promotion problem and the proposed solution. We first formally define the problem:

DEFINITION 1 (CROSS-NETWORK YOUTUBE VIDEO PROMOTION). *Imagine we have a collection of YouTube videos* $\mathcal{V}$ *where each* $v \in \mathcal{V}$ *is represented by its contained textual words and visual keyframes* $[\mathbf{w}_v, \mathbf{f}_v]$*, and a collection of Twitter users* $\mathcal{U}^T$ *whose followees construct the Twitter followee user collection* $\mathcal{U}^{followee} \subset \mathcal{U}^T$*. The goal of Youtube video promotion is: for a given YouTube video* $v \in \mathcal{V}$*, to identify Twitter followee* $u \in \mathcal{U}^{followee}$ *whose followers are most likely to be interested in* v*.*

3.1 Framework

Our solution consists of three stages: *Heterogeneous Topic Modeling*, *Cross-network Topic Association* and *Referrer Identification*(as illustrated in Fig. 2). The goal of Stage 1 is to discover the latent structure within YouTube video and Twitter user spaces, and facilitate the subsequent analysis and applications in topic level. We conduct this by employing generative topic models, with video as *document*, textual word and visual feature of keyframes as the multi-

Table 1: Input (*In*) & output (*Out*) of each stage.

Stage 1	*In*:	YouTube video $v \in \mathcal{V} : [\mathbf{w}_v, \mathbf{f}_v]$; Twitter user $u \in \mathcal{U}^T : \mathcal{U}_u^{followee}$.
	Out:	YouTube video distri. $V : p(\mathbf{z}^Y \vert v)$; Twitter user distri. $U^T : p(\mathbf{z}^T \vert u)$.
Stage 2	*In*:	V, U^T; YouTube, Twitter and overlapped user set $\mathcal{U}^Y, \mathcal{U}^T, \mathcal{U}_o$; YouTube user interested videos $\mathcal{V}_u \subset \mathcal{V}$.
	Out:	Distri. transfer func. $\mathcal{F} : \mathbf{u}^Y \rightarrow \mathbf{u}^T$. ($\mathbf{u}^Y$: the aggregated YouTube user distri.)
Stage 3	*In*:	$\mathcal{F}$; Test video set $\mathcal{V}_t$; Twitter followee set $\mathcal{U}^{followee}$.
	Out:	Twitter followee rank for $v \in \mathcal{V}_t : \psi_v(\cdot)$.

modal *word* in YouTube, and user as *document*, followee as *word* in Twitter. Through this stage, each YouTube video and Twitter user can be represented as distributions in the derived corresponding topic spaces.

As discussed in the introduction, the discrepancy between the cross-network topic spaces prevents from direct analysis. Stage 2 is designed to address this issue by mining the cross-network topic association. Note that traditional semantic-based criteria tend to fail in capturing the association between heterogeneous entities of video and user. We propose a solution that first aggregates YouTube video distribution to user level, and then exploit the overlapped users among different networks as bridge for association mining. The basic premise is that: if the same group of users heavily involve with topic A in network X and topic B in network Y, it is very likely that topic A and B are closely associated. With the derived topic association, topical distribution transfer between different networks is enabled, i.e., given users' topical interest in YouTube videos, we can infer their most probably followed Twitter followee topics.

Since the ultimate goal is to match video to followee. After the offline Stage 1 and Stage 2, in the online Stage 3, we view each test video as a virtual YouTube user who holds identical topical distribution. It is easy to understand that the virtual user actually represents the typical users in YouTube showing significant interest to the test video, who are exactly potential fans and thus the targeted users. Therefore, after topical distribution transfer, it is promising to identify the Twitter followee that best matches the followee topical distribution of the targeted users as the optimal promotion referrer for the video. In Table 1 we summarize the inputs and outputs for each stage.

3.2 Heterogeneous Topic Modeling

3.2.1 YouTube Video Topic Modeling

In YouTube, the video topics are expected to span over both textual and visual spaces. We introduce a modification to the multi-modal topic model, Corr-LDA [22]. Corr-LDA is proposed for the problem of image annotation, by modeling the correspondence between image segments and caption words. It assumes a generative process that first generates the segment descriptions and subsequently the caption words. In our problem, each YouTube video is represented as a pair $(\mathbf{f}; \mathbf{w})$, where $\mathbf{f} = \{f_1, \cdots, f_N\}$ is a collection of N visual feature vectors associated with the extracted keyframes, $\mathbf{w} = \{w_1, \cdots, w_M\}$ is the collection of M caption and tag words. Different from image where each word corresponds to one segment, video caption and tag word usually distribute in several keyframes.

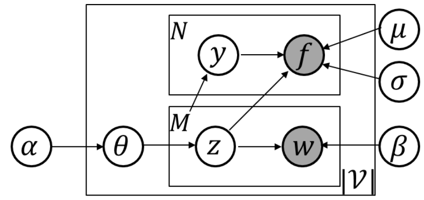

Figure 3: The graphical representation of iCorr-LDA. Note that $\mathbf{y} = \{y_1, \cdots, y_N\}$ are discrete indexing variables that take values from 1 to M with equal probability.

Therefore, we modified the standard CorrLDA and introduce inverse Corr-LDA (iCorr-LDA) to discover the YouTube video multimodal topics. In particular, we first generate M textual words from the standard LDA model. Then, for each of the N keyframes, one of the words is selected and a corresponding keyframe is drawn, conditioned on the same topic generating the word. The graphical model of iCorr-LDA is depicted in Fig. 3. After topic modeling, each video $v \in \mathcal{V}$ can be represented as $\mathbf{v} = \{v_1, \cdots, v_{K^Y}\}$, where K^Y is the number of topics in the derived YouTube video space, $v_k = p(z_k^Y \vert v)$ is video v's topic distribution on the k^{th} topic.

3.2.2 Twitter Followee Topic Modeling

Since the properness of Twitter followee is decided by the followers, we are interested in investigating into the followee-follower architecture in Twitter. Therefore, we represent each Twitter user (*document*) with all his/her followees (*words*) and apply the standard LDA for topic modeling. Since topic modeling exploits co-occurrence relationships, like the YouTube video topics capturing the frequently co-occurred visual features and textual words in videos, the derived Twitter topics actually capture the shared followees by a subset of Twitter users. Particularly, high topic-word distribution indicates the popularity of followees in a group of Twitter followers, and high document-topic distribution indicates user's significant interest in a class of Twitter followees.

After topic modeling, we can obtain Twitter user topic distribution matrix $U^T = \{\mathbf{u}_1^T, \cdots, \mathbf{u}_{\vert\mathcal{U}^T\vert}^T\}$. Each user $u \in \mathcal{U}^T$ is represented as $\mathbf{u}^T = \{u_1^T, \cdots, u_{K^T}^T\}$, where K^T is the number of topics in the derived Twitter followee space, $u_k^T = p(z_k^T \vert u)$ is user u's topic distribution on the k^{th} topic.

3.3 Cross-network Topic Association

3.3.1 YouTube User-Topic Distribution Aggregation

YouTube user's topic distribution can be obtained by aggregating his/her interested videos' distributions. Specifically, for YouTube user u, we construct the interested video set $\mathcal{V}_u \subset \mathcal{V}$ from his/her uploaded videos, favorite videos and videos in the playlists. Given YouTube video $v \in \mathcal{V}_u$ and its topical distribution $p(\mathbf{z}^Y \vert v)$, through simple derivation, we can calculate user u's topical distribution by:

$$p(z_k^Y \vert u_i) = \sum_{v \in \mathcal{V}_u} \frac{N_v(f) + N_v(w)}{N(f) + N(w)} \cdot p(z_k^Y \vert v) \tag{1}$$

where $N_v(f)$, $N_v(w)$ denote the total number of keyframes and words in video v, $N(f) = \sum_{v \in \mathcal{V}_u} N_v(f)$, $N(w) = \sum_{v \in \mathcal{V}_u} N_v(w)$

denote the total number of keyframes and words in video set $\mathcal{V}_u$. After aggregation, we can obtain the YouTube user topic distribution matrix $U^Y = \{\mathbf{u}_1^Y, \cdots, \mathbf{u}_{|\mathcal{U}^Y|}^Y\}$.

3.3.2 Transition Probability-based Association (TP)

With the derived YouTube and Twitter user topic distributions, we present the solutions for topic association mining. Recall that the basic idea is: if many overlapped users who take interests in the i^{th} YouTube topic also follow the j^{th} Twitter topic, the association between the two topics a_{ij} tends to be strong. One direct way is to examine the joint involvement of cross-network topics in the overlapped users.

We assume YouTube and Twitter user set share the overlapped users $\mathcal{U}_o = \mathcal{U}^Y \cap \mathcal{U}^T$. Viewing as a probabilistic transition problem, the topic association can be calculated by aggregating over all the overlapped users [6]:

$$a_{ij} = p(z_j^T|z_i^Y) = \sum_{u\in\mathcal{U}_o} p(z_j^T|u)\cdot p(u|z_i^Y) = \sum_{u\in\mathcal{U}_o} u_j^T \cdot u_i^Y \cdot \frac{p(u)}{p(z_i^Y)}$$

where $p(u)$ is the user prior which we assume having the uniform distribution, the topic prior $p(z_i^Y) = \sum_{u\in\mathcal{U}_o} p(z_i^Y|u)\cdot p(u)$ indicates the popularity of the i^{th} YouTube topic among the overlapped users. By calculating all cross-network topic pairs and subsequent normalization, we can obtain the topic association matrix $A = \{a_{ij}\}_{K^Y\times K^T}$. The distribution transfer from U^Y to U^T can then be fulfilled. Given a new user u_t and the YouTube video topic distribution $p(\mathbf{z}^Y|u_t)$, his/her Twitter followee topic distribution is estimated as:

$$p(z_j^T|u_t) = \sum_{i=1}^{K^Y} a_{ij}\cdot p(z_i^Y|u_t) \tag{2}$$

3.3.3 Regression-based Association

The above probability-based method directly calculates over all overlapped users, where noisy user topic distributions will deteriorate the derived association matrix. An alternative way to obtain the association matrix is to solve an optimization problem. Rewriting the user topic distribution matrices as $U^Y = [U_o^Y, U_{non}^Y]$ and $U^T = [U_o^T, U_{non}^T]$, where U_o^Y, U_o^T denote the overlapped users' distributions on the corresponding topic spaces, we propose to view the association matrix A as the linear regression from the overlapped users' YouTube distribution U_o^Y to their Twitter distribution U_o^T.

Formally, the regression objective function is:

$$\min_A ||U_o^T - AU_o^Y||^2 + \lambda_1||A||_q \tag{3}$$

where the first term represents the regression error, the second term is the regularization penalty used to avoid overfitting, and $\lambda_1 \in [0,1]$ is the weighting parameter. When $q=1$, Eqn. (3) is a lasso problem and can be effectively solved by the feature-sign search algorithm [23]. When $q=2$, Eqn. (3) is a ridge regression problem with analytical solution as:

$$A = U_o^T {U_o^Y}^{\mathrm{T}} (U_o^Y {U_o^Y}^{\mathrm{T}} + \lambda_1 I)^{-1} \tag{4}$$

where $\cdot^{\mathrm{T}}$ is the matrix transpose, and $I \in \mathbb{R}^{K^Y\times K^Y}$ is the identity matrix.

[6] The derivation is based on Bayesian rule, which is omited due to space limitation.

3.3.4 Latent Attribute-based Association (LA)

The aforementioned two association methods are devoted to finding the cross-network association matrix A. Actually, to conduct the topical distribution transfer, the association matrix is not necessarily needed. Moreover, such a matrix exists under the assumption of linear association, which does not hold in complicated cases.

Latent attribute discovery on overlapped users $\mathcal{U}_o^Y, \mathcal{U}_o^T$. (LA_overlap) Instead of pursuing an explicit A for "hard" transfer, we also introduce a third association method, by discovering the shared latent structure behind the two topic spaces. For the overlapped users, the different topic distributions can be viewed as their observed activities on different networks. It is reasonable to assume the latent structure behind these observations is actually user attribute. It is the same user's unique attribute values (e.g., age, gender, occupation, home location, etc.) that give birth to his/her different activities and thus the cross-network topic distributions. In each network, a set of representative topic distribution vectors are extracted as network-specific user factors to represent the latent attributes. Specifically, we assume a YouTube factor $\mathbf{d}^Y = \{d_1^Y, \cdots, d_{K^Y}^Y\}$ and a Twitter factor $\mathbf{d}^T = \{d_1^T, \cdots, d_{K^T}^T\}$ are coupled to the same user attribute $d \in \mathcal{D}$. This can be better understood by analogous to coupled dictionary learning [24]. It is reasonable to assume that the same user should have identical attribute representation, and thus identical coefficients when projected to the coupled user factors.

Formally, let $D^Y = \{\mathbf{d}_1^Y, \cdots, \mathbf{d}_{|\mathcal{D}|}^Y\}$, $D^T = \{\mathbf{d}_1^T, \cdots, \mathbf{d}_{|\mathcal{D}|}^T\}$ denote the coupled user factors in YouTube and Twitter, where $|\mathcal{D}|$ is the number of the latent user attributes. By forcing overlapped user's YouTube and Twitter distributions share the same coefficients after projected to the coupled factors, we have the following optimization objective function:

$$\begin{aligned} \min_{D^Y, D^T, S} \; & ||U_o^Y - D^Y S||_2^2 + ||U_o^T - D^T S||_2^2 + \lambda_2||S||_1 \\ s.t. \; & ||\mathbf{d}^Y||_2^2 \le 1, ||\mathbf{d}^T||_2^2 \le 1, \forall d \in \mathcal{D} \end{aligned} \tag{5}$$

where $S = \{\mathbf{s}_1, ..., \mathbf{s}_{|\mathcal{U}_o|}\}$ with $\mathbf{s}_i$ be the attribute representation for user $u_i \in \mathcal{U}_o$, the constrain $||\mathbf{d}||_2^2 \le 1$ is to prevent D from being arbitrarily large. The reason of using *l1*-norm penalty is to encourage a compact attribute space that users sparsely distribute on. Eqn. (5) can be rewritten as

$$\begin{aligned} \min_{\hat{D}, S} \; & ||\hat{U}_o - \hat{D}S||_2^2 + \lambda_2||S||_1 \\ & s.t. ||\hat{\mathbf{d}}_i||_2^2 \le 1, \forall i \end{aligned} \tag{6}$$

where

$$\hat{U_o} = \begin{bmatrix} U_o^Y \\ U_o^T \end{bmatrix}, \hat{D} = \begin{bmatrix} D^Y \\ D^T \end{bmatrix}$$

The optimization problem (6) can be efficiently solved by the sparse coding algorithm proposed in [23].

Latent attribute discovery on all users $\mathcal{U}^Y, \mathcal{U}^T$. (LA_all) The non-overlapped users have been ignored in the proposed association methods. In practical implementation, plenty of non-overlapped users exist. The optimal user factors should both be coupled to unique latent attributes and well represent the latent structure in each network.

Inspired by this, we reformulate Eqn. (5) that the non-overlapped users $\mathcal{U}_{non}^Y, \mathcal{U}_{non}^T$ also contribute to the user factor discovery in each network, but with no requirement on identical coefficients. Formally, the optimization objective function is:

$$\begin{aligned}\min_{D^Y,D^T,S^Y,S^T} \; & ||U^Y - D^Y S^Y||_2^2 + ||U^T - D^T S^T||_2^2 + \lambda_3 ||S_o||_1 \\ & + \lambda_4 ||S^Y_{non}||_1 + \lambda_5 ||S^T_{non}||_1 \\ s.t. \; & ||\mathbf{d}^Y||_2^2 \le 1, ||\mathbf{d}^T||_2^2 \le 1, \forall d \in \mathcal{D}\end{aligned} \tag{7}$$

where S^Y_{non},S^T_{non} are user factor coefficients for the non-overlapped users in YouTube and Twitter, $S^Y = [S_o, S^Y_{non}]$, $S^T = [S_o, S^T_{non}]$, $\lambda_3, \lambda_4, \lambda_5$ are tuning parameters controlling the factor distribution sparsity. It can be seen that the above formulation learns user factors not only coupled to unique user attributes over the overlapped users, but minimizing the reconstruction error in each network over all the non-overlapped users.

Since Eqn. (7) is convex to $D^Y, D^T, S_o, S^Y_{non}, S^T_{non}$ respectively, we design an iterative algorithm by alternatively optimizing the following three sub-problems till convergence or maximum iteration:

A. Coupled factor distribution learning:

$$\min_{S_o} \; ||U^Y_o - D^Y S_o||_2^2 + ||U^T_o - D^T S_o||_2^2 + \lambda_3 ||S_o||_1 \tag{8}$$

This is exactly the same problem in Eqn. (5) with fixed user factors D^Y, D^T.

B. Divided factor distribution learning:

$$\begin{aligned}&\min_{S^Y_{non}} \; ||U^Y_{non} - D^Y S^Y_{non}||_2^2 + \lambda_4 ||S^Y_{non}||_1 \\ &\min_{S^T_{non}} \; ||U^T_{non} - D^T S^T_{non}||_2^2 + \lambda_5 ||S^T_{non}||_1\end{aligned} \tag{9}$$

This is a multi-task lasso problem and can be solved by the feature-sign search algorithm [23].

C. Coupled user factor update:

$$\begin{aligned}\min_{D^Y,D^T} \; & ||U^Y - D^Y S^Y||_2^2 + ||U^T - D^T S^T||_2^2 \\ s.t. \; & ||\mathbf{d}^Y||_2^2 \le 1, ||\mathbf{d}^T||_2^2 \le 1, \forall d \in \mathcal{D}\end{aligned} \tag{10}$$

This is a quadratically constrained quadratic program problem (QCQP). We utilize an alternative update strategy for solution [25].

With the derived user factors D^Y and D^T, given a new YouTube user topic distribution $\mathbf{u}^Y \in \mathbb{R}^{K^Y \times 1}$, we can estimate the YouTube user factor distribution as:

$$\mathbf{s}^* = \min_s ||\mathbf{u}^Y - D^Y \mathbf{s}||_2^2 + \lambda ||\mathbf{s}||_1 \tag{11}$$

Since unique user shares the same factor coefficients, we can reconstruct his/her Twitter topic distribution as:

$$\mathbf{u}^T = D^T \mathbf{s}^*.$$

3.4 Referrer Identification

With the cross-network distribution transfer function $\mathcal{F}$, we can estimate arbitrary user's Twitter followee topic distribution by inputing his/her YouTube video topic distribution. In our video promotion problem, given a test YouTube video v_t, we simulate a virtual user with identical topic distribution $\mathbf{v}^Y_t = p(\mathbf{z}^Y|v_t)$ to represent the typical YouTube users liking the video [7]. After distribution transfer, the virtual user's Twitter followee topic distribution $\mathbf{v}^T_t = p(\mathbf{z}^T|v_t)$ actually reflects the most probable Twitter following patterns for the video fans.

On the Twitter side, we construct a popular Twitter followee set $\mathcal{U}^{followee}_t \subset \mathcal{U}^{followee}$ serving as the candidate YouTube video promotion referrers. For each popular followee $u \in \mathcal{U}^{followee}_t$, his/her Twitter topic distribution $\mathbf{u}^T$ can be calculated as:

$$p(z^T_k|u) \propto p(u|z^T_k) \cdot p(z^T_k)$$

where $p(z^T_k)$ is the topic prior and can be calculated by aggregating over users. Here $p(z^T_k|u)$ actually reflects followee u's popularity in the k^{th} topic.

Direct product-based matching Given the test YouTube video and candidate Twitter followees represented on the same topic space, one way is to directly use dot product as the properness measure. The properness score of Twitter followee $u \in \mathcal{U}^{followee}_t$ to promote YouTube video v_t is calculated as:

$$properness(u, v_t) = <\mathbf{v}^T_t, \mathbf{u}^T> = \sum_{k=1}^{K^T} v^T_{t,k} \cdot u^T_k \tag{12}$$

A rank $\psi_{v_t}(\cdot)$ defined on the followees can be obtained accordingly to identify the optimal Twitter referrer.

Weighted product-based matching We also investigate a matching strategy by optimizing the weights for each topic. Viewing test video as the query and candidate Twitter followee set as the collection, the referrer identification can be treated as a retrieval problem. In light of this, we design a training scheme and adopt ranking SVM [26] for topic selection.

Ranking SVM model is with the form as:

$$g(\cdot, \cdot) = \mathbf{h} \cdot \phi(\cdot, \cdot) \tag{13}$$

where $\mathbf{h}$ is the model parameter, i.e., the learnt weights for the corresponding topics. The goal of ranking SVM is to learn an optimal $\mathbf{h}$ that best maintains the rank order in the training query-document pairs.

In our problem, we define the feature mapping function as the vector product between video and followee distributions:

$$\phi(\mathbf{v}^T_t, \mathbf{u}^T) = \mathbf{v}^T_t \odot \mathbf{u}^T = \{v^T_{t,1} \cdot u^T_1, ..., v^T_{t,K^T} \cdot u^T_{K^T}\}$$

where $\odot$ indicates the element-wise multiplication. To obtain the ranks in the training set, for each query-document pair v, u, we need to calculate their ground-truth properness score. According to the discussion in introduction, the properness of Twitter followee is decided by how many of his/her followers like the test video. Therefore, we combine two information retrieval metrics of *precision* and *recall* to define the Ground-Truth (GT) properness:

$$\text{precision}(v, u) = \frac{|\mathcal{U}_v \cap \mathcal{U}^{followee}_u|}{|\mathcal{U}^{followee}_u|}$$

$$\text{recall}(v, u) = \frac{|\mathcal{U}_v \cap \mathcal{U}^{followee}_u|}{|\mathcal{U}_v|}$$

$$\text{GT-properness}(v, u) = \frac{2}{\text{precision}(v, u)^{-1} + \text{recall}(v, u)^{-1}}$$

[7] Due to the flexibility of iCorr-LDA, we can also estimate the topic distribution for videos with only visual keyframes or textual words. This extends the applicability of our framework.

Table 2: Statistics of our dataset.

$\lvert\mathcal{U}^Y\rvert$	$\lvert\mathcal{U}^T\rvert$	$\lvert\mathcal{U}_o\rvert$	$\lvert\mathcal{V}\rvert$	Avg.$\lvert\mathcal{U}_u^{followee}\rvert$
38,540	39,400	11,850	2,280,129	891.1

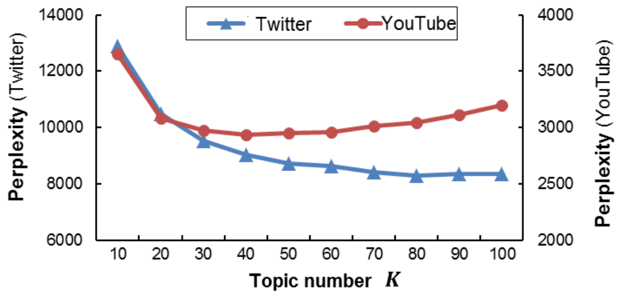

Figure 4: The perplexities for different topic numbers on YouTube and Twitter.

where $\mathcal{U}_v$ is the set of users showing interest in v, $\mathcal{U}_u^{followee}$ is the follower set of u [8]. We can see that *recall* actually concerns with coverage of the interested YouTube audiences, while *precision* is in charge of the virtual cost. With the learnt $\mathbf{h}^*$, the properness of Twitter followee $u \in \mathcal{U}_t^{followee}$ for test video v_t is calculated as:

$$properness(u, v_t) = \mathbf{h}^* \cdot \phi(\mathbf{v}_t^T, \mathbf{u}^T). \tag{14}$$

4. EXPERIMENTS

4.1 Dataset

Since no ready cross-network dataset is available, we construct a new dataset with user account linkage between YouTube and Twitter. Google+ encourages users to provide the external links to their other social media network accounts. We first collected 143,259 Google+ users, among which 38,540 provide YouTube account, 39,400 provide Twitter account, 11,850 provide both accounts [9]. For each YouTube user, we further downloaded his/her uploaded videos, favorite videos, playlists and the involved video information via YouTube API. For each Twitter user, we downloaded his/her followee set and user profiles via Twitter API. Table 2 summarizes the key statistics [10].

4.2 Heterogeneous Topic Modeling

4.2.1 Topic Number Selection

In topic modeling, the selection of topic number is very important. We resort to the perplexity in this paper, which is a standard measure for estimating how well one generative model fits the data [28]. The lower the perplexity score is, the better the performance. We test the perplexity with different topic number K^Y and K^T on 490,000 held-out YouTube videos and 9,400 held-out Twitter users, respectively [11]. The perplexity scores on different topic numbers are shown in Fig. 4. We can see that on both YouTube and Twitter, the perplexities decrease dramatically first before reaching a relatively stable level and then have a tendency to increase when the models are overfit. Since larger topic number requires more computational cost and has overfitting risk, we prefer the smallest topic number that leads to perplexity on the stable level. Therefore, we choose the topic number $K^Y = 40$ for YouTube and $K^T = 80$ for Twitter.

Table 3: Visualization of discovered YouTube topics.

Topic #1	Word	gameplay xbox playstation gaming minecraft
	Video	"Epic Mods - MW2 MOD IN CoD4" 
		"HEXXIT COOP ep7 w/ Double" 
		"Halo 4 Adrift Multiplayer Map" 
Topic #17	Word	history german berlin germany poetry
	Video	"GEH STERBEN, DU OPFER!!!" 
		"Syrien - Wahrheit ĺźber das Massaker" 
		"Volker Pispers - Einzeltater" 

4.2.2 Visualization of Discovered Topics

In order to interpret the derived topic spaces, we visualize some of the discovered topics in YouTube and Twitter, respectively. Table 3 shows two sampled YouTube video topics. For each topic, we provide the top-5 probable words and 3 most representative videos. Representative videos are ranked based on the video-topic distribution $p(z_k^Y|v)$ and represented by the keyframes and video titles in Table 3. By visualizing both the semantic and visual information, it is very easy to interpret the domain knowledge associated with each topic. Moreover, the discovered video topics show high consistency between textual semantics and visual patterns.

Table 4 shows three sampled Twitter followee topics, with each visualized by the top-3 probable followees and the followees' profile information. It is conceived that the discovered Twitter topics have a quite wide coverage: the general topic #43 addressing the game-related popular followees, the specific topic #10 consisting of Forbes influencers, and even the geographic topic #38 with the top followees all coming from Berlin. Twitter users' joint following patterns are well captured in modeling the follower-followee relationship, which is very important to the subsequent promotion application.

[8] For a given YouTube video, we can only know whether a specific Twitter followee's followers will like the video if we know these Twitter followers' YouTube accounts, so both the sets $\mathcal{U}_v$ and $\mathcal{U}_u^{followee}$ are counted based on the known overlapped user set $\mathcal{U}_o$.

[9] User linkage mining is a separate topic in cross-network analysis [27]. In our work, to guarantee a promising overlapped user resource, we leverage user self-provided account links on Google+.

[10] Avg.$\lvert\mathcal{U}_u^{followee}\rvert$ is the average number of followees over all the examined Twitter users.

[11] Hyperparameters are fixed as $\alpha = 0.8$ and $\beta = 0.1$ according to the empirical expectation for the output distribution [29].

Table 4: Visualization of discovered Twitter topics.

Topic	User ID	Username	Location	#followers	Self-description
#43	63485337	Markus Persson	Stockholm, Sweden	1,436,534	Hey, you! Play more games! Now!
	36803580	Steam	–	932,044	Steam, The Ultimate Online Game Platform. Follow us...
	11167502	Humble Bundle	San Francisco,CA	192,764	News from the Humble Bundle. For support, please...
#10	21279340	Pam Moore	Orlando, FL	178,101	50% mktg 50% geek CEO, Forbes TOP Social Influencer.
	33057154	Jeff Sheehan	Atlanta, GA	254,984	Social Media Pro | Speaker | Author | 30+ years Mktg.
	15081182	Warren Whitlock	Las Vegas, NV	178,759	Forbes Power Influencer. Radio Host, Author, Speaker...
#38	5876652	Sascha Lobo	Berlin, Germany	161,099	Author, Internet.
	9655032	netzpolitik	Berlin, Germany	120,014	Entrepreneur, activist, organizer of @republica.
	9334352	Mario Sixtus	Berlin, Germany	60,542	Journalist, Photographer. Hier mehr oder weniger

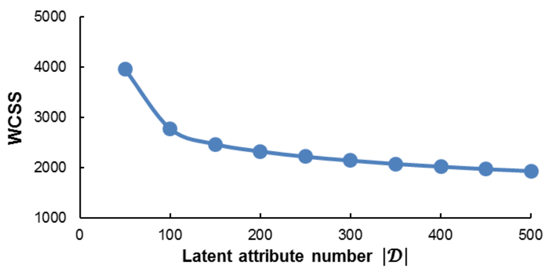

Figure 5: Sum of WCSS over different $|\mathcal{D}|$.

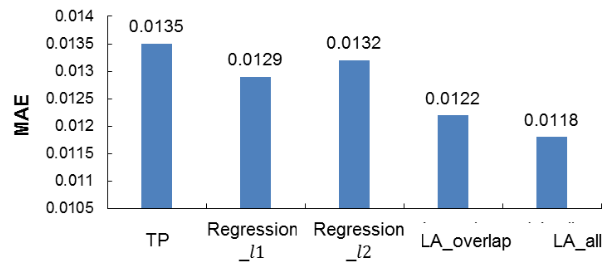

Figure 6: MAE for distribution transfer in Stage 2.

4.3 Cross Network Topic Association

4.3.1 Experimental Setting

Given a user with his/her YouTube topic distribution $\mathbf{u}^Y$, the goal of Stage 2 is to estimate his/her Twitter followee topic distribution $\mathbf{u}^T$. Therefore, we utilize Mean Abosolute Error (MAE) as the evaluation metric. We randomly select half of the overlapped users to construct the test set $\mathcal{U}_{test}$, and the rest overlapped users and non-overlapped users as the training set. MAE is calculated over all topics of each test user as:

$$\text{MAE} = \frac{\sum\limits_{u \in \mathcal{U}_{test}} \sum\limits_{k=1}^{K^T} |\hat{u}_k^T - u_k^T|}{|\mathcal{U}_{test}| K^T} \tag{15}$$

where u_k^T and $\hat{u}_k^T$ are the actual and estimated user u's topic distribution on the k^{th} Twitter topic.

For model parameters, we select the regularization coefficient λ_1 in Eqn. (3) by grid search and 5-fold cross validation. Tuning parameters λ_2 in Eqn. (5) and $\lambda_3, \lambda_4, \lambda_5$ in Eqn. (7) are selected by a combined line-search strategy according to the minimal objective energy after converge. As a result, the parameters are set as $\lambda_1 = 0.1, \lambda_2 = 0.2, \lambda_3 = 0.1, \lambda_4 = 0.01, \lambda_5 = 0.01$. It is particularly non-trivial to decide the number of attributes $|\mathcal{D}|$ in the latent attribute-based association methods: small $|\mathcal{D}|$ may fail to capture the intrinsic structures, while big $|\mathcal{D}|$ will lead to overfitting. As discussed in the solution section, the coupled user factors can be understood as a pair of dictionaries in the discovered Twitter and YouTube topic spaces. Classical clustering metrics, e.g., Within-Cluster Sum of Squares (WCSS) [30], are widely used to evaluate how well the observed data can be reconstructed from the learnt dictionary. Therefore, we conduct K-means on YouTube and Twitter user distributions U^Y, U^T with the identical cluster number $|\mathcal{D}|$. In Fig. 5 we draw the curve of WCSS sum on the two networks w.r.t. the change of $|\mathcal{D}|$. We choose $|\mathcal{D}| = 300$ when the aggregated reconstruction error decreases to a steadily low level.

4.3.2 Experimental Results and Analysis

The transition probability (TP) and regression based methods all yield explicit topic association matrix. To better understand the association between heterogeneous topics, we first examine the derived association matrix $A \subset \mathbb{R}^{K^Y \times K^T}$ from TP. Among the $K^Y \times K^T = 3,200$ topic association pairs, the most significant are $\{z_1^Y, z_{43}^T\}$ and $\{z_{17}^Y, z_{38}^T\}$, which have been visualized in Table 3 and 4. We can see that the derived association involves with multiple aspects: game-related YouTube topic #1 significantly associates with Twitter topic #43 whose top-ranked followees are official game platforms or developers, and the association between YouTube topic #17 and Twitter topic #38 results from their shared location in Germany. A single association metric, e.g., semantics, tends to fail in this case. Actually, one advantage of exploiting the overlapped users for association mining is its flexibility: there is no need to explicitly design an association metric, and users' collaborative activities on different social networks define the metric.

Performance comparison among the proposed methods is shown in Fig. 6. Several observations are made. (1) With MAE lower than 0.015, all the proposed association solutions achieve satisfied performance. This shows the reasonability by exploiting the overlapped users towards cross-network topic association. (2) The latent attribute-based methods (*LA_overlap*, *LA_all*) outperform the explicit association matrix-oriented methods (*TP*, *Regression*). In addition to the freedom to non-linear association, *LA*-based solutions address the hidden structure behind the observed heterogeneous user activities and enjoy better interpretation. (3) By considering the non-overlapped users, *LA_all* is slightly superior to *LA_overlap*. This validates our assumption that better capturing the latent structure in each network contributes to improved coupled factor discovery.

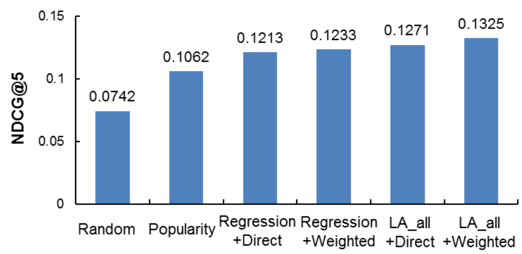

Figure 7: NDCG@5 for different settings in Stage 3.

4.4 Twitter Referrer Identification

4.4.1 Experimental Setting

2,061 videos that more than 15 overlapped users have shown interest to are selected to construct the YouTube test video set $\mathcal{V}_t$. Meanwhile, 21,276 Twitter followees who are followed by more than 50 users construct the candidate Twitter followee set $\mathcal{U}_t^{followee}$.

We use Normalized Discounted Cumulative Gain (NDCG) as the evaluation metric, which is widely used in retrieval problems. NDCG is defined as:

$$NDCG@k = \frac{1}{Z}\sum_{j=1}^{k}\frac{2^{rel(j)}-1}{\log(1+j)} \tag{16}$$

where $rel(\cdot)$ is a relevance function between the test video and the ranked followee candidate. With the goal to identify Twitter followees with optimal coverage-cost balance, we use *GT-properness* as in Eqn. (14) to calculate $rel(\cdot)$. Moreover, the groundtruth for the Twitter referrer ranking is also constructed by sorting the *GT-properness* between the test video and the followee candidate based on the known overlapped user set $\mathcal{U}_o$.

We consider the following settings for comparison:

- *Random*: randomly select k followees from $\mathcal{U}_t^{followee}$;
- *Popularity*: select k popular Twitter followees with the most #followers;
- *Regression+Direct*: distribution transfer by *Regression_l*1, matching by *Direct product*;
- *Regression+Weighted*: distribution transfer by *Regression_l*1, matching by *Weighted product*;
- *LA_all+Direct*: distribution transfer by *LA_all*, matching by *Direct product*;
- *LA_all+Weighted*: distribution transfer by *LA_all*, matching by *Weighted product*.

4.4.2 Experimental Results and Analysis

We show NDCG@5 for different settings in Fig. 7 [12]. It is observed that: (1) *Popularity* fails to identify the optimal Twitter referrer. This is easy to understand. While high #follower guarantees the coverage of potential viewers (*precision*), the retrieved follower set is expected to also include many uninterested users (*recall*), which deviates our goal towards target promotion. (2) Conducting distribution transfer by *LA_all+Direct* and *LA_all+Weighted* obtain better performance than *Regression+Direct* and *Regression+Weighted*. This coincides with our motivation that more accurate distribution transfer contributes to improved referrer identification. (3) The settings with weighted product-based matching consistently outperform those with direct product. This demonstrates the advantage of topic weight optimization. One possible interpretation is that different topics contribute differently in view of referrer identification.

[12] Since the view history of Youtube users is unavailable via public API, the NDCG results are relatively low due to a lack of abundant user behaviors.

5. DISCUSSION

5.1 Application and Extension

Application. The proposed framework also enables solutions to other applications. From Stage 2, we actually obtain the association between YouTube video interests and Twitter following patterns. Based on this association, cross-network personalized recommendation problems on two directions can be enabled: recommending Twitter followee topic or Twitter lists given YouTube video interest [16], and recommending YouTube videos given Twitter followee list.

Another promising application is on examining the value of Twitter followees. Current methods value Twitter followees by directly analyzing their followers' demographics information, e.g., the followee has a lot of young female followers. The proposed framework in this work facilitates application-oriented Twitter followee value analysis, by associating Twitter followee topic space with the needed topic spaces. For example, our work can be viewed as valuing Twitter followee w.r.t. promotion efficiency to YouTube videos. This significantly expands understanding into the value of Twitter followees.

Extension. Our current solution only employs the content feature of YouTube test videos, i.e., title, tags and keyframes. One extension is to combine with social features, e.g., who uploads or favorites the video. The consideration of user social network is also expected to contribute to improved cross-network association.

Moreover, the current referrer identification is on the individual level, i.e., no interaction between followees is considered. In practice, when choosing a group of followees as the promotion referrers, follower intersection of the candidate followees need to be modeled. Analogous to advertising, as discussed in the introduction, this work actually addresses the problem of advertising media selection. Other problems in advertising include advertising anchor text generation (i.e., optimizing video description for promotion), and advertising slot bid (i.e., followee reshare time selection).

Other than crawling user activities from public API, another way for evaluation is to run a real-world system and ask the users for participation. However, the examined users are popular Twitter followees, who are costly or completely unwilling to participate into the study. Therefore in the future, we are expecting to conduct a real-world evaluation by examining some less popular followees instead.

5.2 A Promising Direction

The idea of exploiting overlapped users towards cross-network association actually opens up possibilities to a very interesting direction. People involve with social media by interacting with heterogeneous social media knowledge, e.g., multimedia semantics, geographic patterns, people consuming patterns, social interactions, etc. The association among different social media activities will lead to insightful observations, contribute to collective utilization, and facilitate advanced social media analysis and applications. For example, the association between user watching activity in *YouTube* and transaction activity in *Amazon* leads to understanding between user interest and consuming models, and facilitates cross-network product target advertising. "Multimedia" research under social media circumstances may understand not only text, image, video, but the association among heterogeneous social media knowledge.

The user-generated nature of social media inspires us to understand the heterogeneous knowledge by "how we experience the world". Instead of conducting analysis from scratches, the different activities that overlapped users contribute in different social media networks can be employed as human supervision. This actually borrows the essence of crowdsourcing where the collective human intelligence is aggregated.

The guideline to instantiate the idea is: (1) Determine the heterogeneous knowledge involved in different social media networks, and crawl a dataset of overlapped users and their heterogeneous activities. (2) Extract the latent topic spaces on each network to construct heterogeneous knowledge bases. (3) Conduct cross-network topic association by exploiting the observed overlapped users as supervision. (4) Design cross-network collaborative applications (one-way or two-way) based on the derived knowledge association.

6. CONCLUSION

We have proposed an overlapped user-based association solution framework, to address the novel cross-network YouTube video promotion problem. Alternative methods have been developed and evaluated, to demonstrate the effectiveness of exploiting user collaboration towards heterogeneous knowledge association. The proposed framework is quite flexible, and can be generalized to other cross-network collaborative problems. We hope that this paper could serve as a good chance to emphasize the collective utilization of social media sources and further the agenda of cross-network analysis and application in social multimedia research.

7. ACKNOWLEDGMENT

This work is supported in part by National Basic Research Program of China (No. 2012CB316304), National Natural Science Foundation of China (No. 61225009, 61332016, 61303176), and Beijing Natural Science Foundation (No. 4131004).

8. REFERENCES

[1] Bernardo A Huberman, Daniel M Romero, and Fang Wu. Crowdsourcing, attention and productivity. *Journal of Information Science*, 35(6):758–765, 2009.

[2] Jean Burgess and Joshua Green. *YouTube: Online video and participatory culture.* John Wiley & Sons, 2013.

[3] Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon. I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In *IMC 2007*, pages 1–14.

[4] Xu Cheng, Cameron Dale, and Jiangchuan Liu. Statistics and social network of youtube videos. In *IWQoS 2008*, pages 229–238.

[5] Renjie Zhou, Samamon Khemmarat, and Lixin Gao. The impact of youtube recommendation system on video views. In *IMC 2010*, pages 404–410.

[6] Youmna Borghol, Sebastien Ardon, Niklas Carlsson, Derek Eager, and Anirban Mahanti. The untold story of the clones: Content-agnostic factors that impact youtube video popularity. In *KDD 2012*, pages 1186–1194.

[7] Flavio Figueiredo, Jussara M Almeida, Marcos André Gonçalves, and Fabrício Benevenuto. On the dynamics of social media popularity: A youtube case study. *arXiv preprint arXiv:1402.1777*, 2014.

[8] Delip Rao, David Yarowsky, Abhishek Shreevats, and Manaswi Gupta. Classifying latent user attributes in twitter. In *SMUC 2010*, pages 37–44.

[9] Zhiheng Xu, Rong Lu, Liang Xiang, and Qing Yang. Discovering user interest on twitter with a modified author-topic model. In *WI-IAT 2011*, pages 422–429.

[10] John R. Kender, Matthew L. Hill, Apostol (Paul) Natsev, John R. Smith, and Lexing Xie. Video genetics: A case study from youtube. In *ACM Multimedia 2010*, pages 1253–1258.

[11] Fabian Abel, Samur Araújo, Qi Gao, and Geert-Jan Houben. Analyzing cross-system user modeling on the social web. In *Web Engineering*, pages 28–43. Springer, 2011.

[12] Zhengyu Deng, Jitao Sang, and Changsheng Xu. Personalized video recommendation based on cross-platform user modeling. In *ICME 2013*, pages 1–6. IEEE.

[13] Suman Deb Roy, Tao Mei, Wenjun Zeng, and Shipeng Li. Socialtransfer: cross-domain transfer learning from social streams for media applications. In *ACM Multimedia 2012*, pages 649–658. ACM.

[14] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and P Krishna Gummadi. Measuring user influence in twitter: The million follower fallacy. *ICWSM*, 10:10–17, 2010.

[15] Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. Twitterrank: finding topic-sensitive influential twitterers. In *WSDM 2010*, pages 261–270. ACM.

[16] Saptarshi Ghosh, Naveen Sharma, Fabricio Benevenuto, Niloy Ganguly, and Krishna Gummadi. Cognos: crowdsourcing search for topic experts in microblogs. In *SIGIR 2012*, pages 575–590. ACM.

[17] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *KDD 2003*, pages 137–146. ACM.

[18] Wei Chen, Chi Wang, and Yajun Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *KDD 2010*, pages 1029–1038. ACM.

[19] Brendan F Klare, Zhifeng Li, and Anil K Jain. Matching forensic sketches to mug shot photos. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(3):639–646, 2011.

[20] Zhen Lei, Matti Pietikainen, and S Li. Learning discriminant face descriptor. 2013.

[21] David R Hardoon, Sandor Szedmak, and John Shawe-Taylor. Canonical correlation analysis: An overview with application to learning methods. *Neural computation*, 16(12):2639–2664, 2004.

[22] David M Blei and Michael I Jordan. Modeling annotated data. In *SIGIR 2003*, pages 127–134.

[23] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y Ng. Efficient sparse coding algorithms. *Advances in neural information processing systems*, 19:801, 2007.

[24] Jianchao Yang, Zhaowen Wang, Zhe Lin, Scott Cohen, and Thomas Huang. Coupled dictionary training for image super-resolution. *Image Processing, IEEE Transactions on*, 21(8):3467–3478, 2012.

[25] Meng Yang, Lei Zhang, Jian Yang, and David Zhang. Metaface learning for sparse representation based face recognition. In *ICIP 2010*, pages 1601–1604.

[26] Thorsten Joachims. Optimizing search engines using clickthrough data. In *KDD 2002*, pages 133–142. ACM.

[27] Jing Liu, Fan Zhang, Xinying Song, Young-In Song, Chin-Yew Lin, and Hsiao-Wuen Hon. What's in a name?: an unsupervised approach to link users across communities. In *WWW*, pages 495–504, 2013.

[28] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

[29] Thomas L Griffiths and Mark Steyvers. Finding scientific topics. *Proceedings of the National academy of Sciences of the United States of America*, 101(Suppl 1):5228–5235, 2004.

[30] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. *Applied statistics*, pages 100–108, 1979.