YouTube Video Promotion by Cross-network Association: @Britney to Advertise gangnam style

Ming Yan, Jitao Sang, Changsheng Xu, Fellow, IEEE, and M. Shamim Hossain, Senior Member, IEEE

Abstract

The emergence and rapid proliferation of various social media networks have reshaped the way how video contents are generated, distributed and consumed in traditional video sharing portals. Nowadays, online videos can be accessed from far beyond the internal mechanisms of the video sharing portals, such as internal search and front page highlight. Recent studies have found that external referrers, such as external search engines and other social media websites, arise to be the new and important portals to lead users to online videos. In this paper, we introduce a novel cross-network collaborative application to help drive the online traffic for given videos in traditional video portal YouTube by leveraging the high propagation efficiency of the popular Twitter followees. Since YouTube videos and Twitter followees
distribute on heterogeneous spaces, we present a cross-network association-based solution framework. In this framework, we first represent YouTube videos and Twitter followees in the corresponding topic spaces separately by employing generative topic models. Then, the cross-network topic spaces are associated from both semantic-based and network-based perspectives through the collective intelligence of the observed overlapped users. Based on the derived cross-network association, we finally match the query YouTube videos and candidate Twitter followees in the same topic space with a unified ranking method. The experiments on a real-world large-scale dataset of more than 2.2 million YouTube videos and 31.8 million tweets from 38,540 YouTube users and 39,400 Twitter users demonstrate the effectiveness and superiority of our solution in which network-based and semantic-based association are integrated.

Index Terms

video promotion, cross-network analysis, social media

I. INTRODUCTION

With the rise of social media, the way people can get access to the video contents is changing. Instead of only relying on the internal mechanisms provided by the traditional video sharing portal to access the videos, more and more people now prefer to directly watch videos from their involved social media networks [12]. ForeSee has reported that more than 18% users are influenced by the social network when accessing video contents\(^1\) and a significant portion of what users watch is being increasingly referred by social media [33]. Some recent research work has also begun to investigate into the interaction between video sharing and social media networks [9], [24]. Social media brings in much possibility to the traditional video sharing portals and it is very fascinating to explore the innovation sparkles generated when these two meet with each other.

Latest statistics show that, for the world’s largest video sharing portal YouTube, 100 hours of videos are uploaded within every minute which results in an estimate of more than 2 billion videos totally \(^2\). However, in spite of the massive videos generated in YouTube, it exhibits limited propagation efficiency with the internal video-centric mechanisms and many high quality videos may remain unknown to the wide public. According to research, YouTube video view count distribution exhibits a power-law pattern with truncated tails [6]. 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 [8]. On the other hand,

\(^1\) http://www.foreseeresults.com.

thefollowee-follower architecture and fast diffusion characteristic of the social microblogging service Twitter [21] has established itself as a great external platform to promote and engage with the audiences and distinguished itself with significant information propagation efficiency. In this paper, we focus on how to utilize the auxiliary Twitter social network to help promote videos in traditional video sharing portal YouTube.

Recently, it has been reported that over 700 YouTube videos are shared in Twitter each minute [28] and Twitter has allowed 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. Under this followee-follower structure, 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 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).

Three challenges are mainly concerned with our problem: (1) In our scenario, 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, the Twitter followee and YouTube video distribute in completely different
social platforms and no explicit association exists between them; (2) We can only know the followers’ activities on Twitter, based on what only the the demographics or interests on the general level can be inferred [27], [34]. While, the YouTube videos are known to distribute more on specific semantic level [19]. The discrepancy in topic granularity makes it impractical to directly evaluate Twitter followers’ interest to YouTube videos; (3) Although more popular Twitter followees (with more followers) can result in higher coverage to the general 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. Besides, the more popular the Twitter followee is, the more cost is needed to get him/her to help. Therefore, both the audience coverage and virtual cost should be considered when measuring the “properness” of Twitter followees for specific YouTube videos.

To address the above challenges, we design a coherent three-stage solution framework (as depicted in Fig. 2). First, since a practical solution should be not limited to the specific video and followee but generalizable on the alike sets, we propose to investigate the problem on some generalized topic level. Afterwards, different kinds of cross-network association between YouTube and Twitter are established separately on the corresponding topic spaces. Based on the learnt cross-network association, we finally propose to match the Twitter followees with the target YouTube video in a supervised ranking procedure with both audience coverage and virtual cost considered.

In our solution, the key lies in how to establish a reasonable cross-network association between YouTube and Twitter. Inspired by the fact that the same individual usually involves with different social media networks and different social media networks share remarkable percentage of overlapped users ³, 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. Therefore, we propose a brand new way to establish the cross-network association by leveraging the collective intelligence of the observed overlapped users.

Since YouTube video generally distributes on specific semantic level, a direct way to associate the YouTube and Twitter spaces is from the content level where the YouTube video content and Twitter users’ generated tweets are used to capture the association. However, this kind of association can only capture the semantic correlation between the two spaces and still suffers from the discrepancy issue in topic granularity. Therefore, we also make further exploration with a more flexible and heterogeneous kind of association under which the YouTube video content and the network structure of the Twitter users

³ See “Anderson Analytics 2009 report: what your favorite social network says about you?”.
are correlated. Different perspectives of correlations between YouTube and Twitter spaces can be obtained with these semantic-based and network-based association methods complemented with each other. By considering both of the cross-network associations in a combined ranking procedure, we finally achieve better ranking performance to identify the proper Twitter followees to promote the target videos.

Our contributions in this work can be summarized as follows:

- Problem-wise, 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.
- Solution-wise, a cross-network association-based solution framework is presented, under which different kinds of associations are explored and integrated. Alternative methods have been examined and we have introduced a novel and promising perspective by exploiting the overlapped users as bridge to obtain the association (which is inspired by crowd wisdom). To the best of our knowledge, this is the first attempt to mine the cross-network association under a user-bridged scheme.

A preliminary version of this work was published in [35]. The extension in this paper includes three aspects: (1) Except for the network-based association in [35], we also add a new semantic-based association parallelly by leveraging users’ generated tweet information. Each stage for this new association is also designed and evaluated separately; (2) We explicitly present a unified ranking framework to integrate both semantic-based association and network-based association for better Twitter referrer identification in Section III-E; (3) More quantitative experiments are conducted to validate the flexibility and effectiveness of our proposed solution framework, including: (a) A direct Semantic Match method is added as a baseline in Section IV-D2; (b) A new intuitive user coverage (Cov) metric is added to evaluate the performance of different methods in terms of the actual size of users who will adopt the target videos in Section IV-D2 and IV-D3; (c) A more focused comparison and analysis between single semantic-based association and network-based association are made in Section IV-D3.

The remaining of this paper is organized as follows: Section II reviews the related work. Section III formulates the cross-network video promotion problem and elaborates the proposed three-stage solution framework. Experimental results are reported in Section IV. Finally, we conclude the paper in Section VI.

II. RELATED WORK

This section reviews the related topics. Instead of a complete coverage, we only review the representative works in each topic, with the goal to position this work in the coordinate of existing works for
better understanding the addressed problem and the proposed solution.

A. 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 [1], 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 [10]. 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 [30]. In [26], Twitter event detection is conducted by employing Wikipedia pages as the authoritative references. 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.
B. Social Media Influencer Mining

Previous analysis on Twitter has found that popular users with high in-degree are not necessarily influencers for propagation [5], which calls for research onto the problem of 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 [32], and incorporating auxiliary sources like Twitter lists [13]. Another line is concerned with maximizing influence spread by initializing some seed users. David et al. first defined this problem [18], which is then applied to product adoption [2] and viral marketing [7].

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.

C. 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 heterogeneous face recognition and cross-media retrieval, 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 [20]. In [23], 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, among which subspace clustering has proved to be a very effective method to represent data from multiple sources and modalities [31]. In [11], Elhamifar et al. introduced a sparse subspace clustering method, which demonstrates its effectiveness on the two real-world problems of motion segmentation and face clustering. Roy et al. further presented a dynamic heuristic subspace clustering algorithm for weather prediction [29]. Multimodal topic modeling can also be viewed as one type of subspace learning, where multimodal representations are projected to a shared topic space [3].
Fig. 3. Association mining by leveraging the overlapped users in Stage 2: both the semantic-based association and network-based association are obtained under this same scheme. Since four users are both involved with topic $z^T_2$ in Twitter and $z^Y_2$ in YouTube, $z^T_2$ and $z^Y_2$ are closely associated under our scheme that they may be both related with music. The topic is indicated by the top-3 followee/tweet word or video tag in terms of the topic-word distribution in Stage 1.

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.

III. CROSS-NETWORK YOUTUBE VIDEO PROMOTION

In this section, we first formally define the problem. Then, the overview of our solution framework is given. Finally, we introduce each part of the framework in detail.

A. Problem Definition

**Definition 1. Cross-network YouTube video promotion.** Given a collection of YouTube videos $\mathcal{V}$ where each video $v \in \mathcal{V}$ is represented by its contained textual words and visual keyframes $[w_v, f_v]$, and a collection of Twitter users $\mathcal{U}^T$ where each user $u \in \mathcal{U}^T$ is represented by his/her followee collection and
generated tweet collection $[U_{u_{\text{followee}}}^t, U_{u_{\text{tweet}}}^t]$, the followees of these Twitter users construct the Twitter followee user collection $U_{\text{followee}} \subset U^T$ and part of the corresponding YouTube user accounts are also known. The goal of YouTube video promotion is: for a given YouTube video $v \in V$, to identify Twitter followee $u \in U_{\text{followee}}$ whose followers are most likely to be interested in $v$.  

B. 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 multimodal word in YouTube. In Twitter, a semantic-based and a network-based topic spaces are constructed simultaneously via topic models, with user as document, tweet word or followee as word, respectively. Through this stage, each YouTube video and Twitter user
can be represented as distributions in the derived corresponding topic spaces.^^4^^

As discussed in the introduction, no explicit association 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. To obtain a flexible association that goes beyond the traditional semantic-based criteria, we propose a user-contributed solution that first aggregates YouTube video distribution to user level, and then exploits the overlapped users among different networks as bridge for association mining (as depicted in Fig. 3). 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. To address the topic discrepancy issue between cross-network topic spaces, both semantic-based association and network-based association are conducted separately under this scheme. 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 (1) for semantic-based association: their most favorite Twitter tweet semantic topics; (2) for network-based association: 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 both the Twitter tweet semantic and Twitter followee topical distributions of the targeted users as the optimal promotion referrer for the video (as illustrated in Fig. 4). In Table I we summarize the inputs and outputs for each stage.

C. Heterogeneous Topic Modeling

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 [3]. 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

^^4^^ In practical implementation, consideration of topic evolution is also important since Twitter streams and YouTube videos both change over time. In this paper, we only focus on the static user interest and leave the exploration of temporal dynamics as the future work.
Fig. 4. A brief flowchart for the referrer identification in Stage 3 by integrating both semantic-based association and network-based association in a unified matching scheme.

A pair \((f; w)\), where \(f = \{f_1, \cdots, f_N\}\) is a collection of \(N\) visual feature vectors associated with the extracted keyframes, \(w = \{w_1, \cdots, w_M\}\) 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.

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. 5.

After topic modeling, each video \(v \in V\) can be represented as \(v = \{v_1, \cdots, v_{K^Y}\}\), where \(K^Y\) is the number of topics in the derived YouTube video space, \(v_k = p(z_k^Y | v)\) is video \(v\)’s topic distribution on the \(k^{th}\) topic.

2) Twitter User Topic Modeling: In Twitter, a semantic-based and a network-based topic spaces on user level are constructed simultaneously via standard topic models.
**Semantic-based Twitter User Topic Modeling**  Since YouTube videos distribute more on semantic level, it is natural to also represent Twitter users in some semantic topic space. Therefore, we aggregate each Twitter user’s generated tweets and keep only the nouns and hashtags\(^5\). Then the standard LDA model is applied to each user for topic modeling, with user as *document*, the nouns or hashtags of his/her tweets as *word*. In this way, the derived Twitter topics can capture some co-occurred semantic concepts frequently used by many users which may also be found in YouTube video semantic space.

**Network-based Twitter User Topic Modeling**  Since the properness of Twitter followee is decided by the followers, we are also 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 (1) Twitter user tweet topic distribution matrix \(U^T_t = \{u^T_t1, \ldots, u^T_tK^T_t\}\); (2) Twitter user followee topic distribution matrix \(U^T_f = \{u^T_f1, \ldots, u^T_fK^T_f\}\). Each user \(u \in \mathcal{U}^T\) is represented as \(u^T_t = \{u^T_t1, \ldots, u^T_{K^T_t}\}\) and \(u^T_f = \{u^T_f1, \ldots, u^T_{K^T_f}\}\) in the corresponding topic spaces, where \(K^T_t\) and \(K^T_f\) are the number of topics in the derived Twitter topic spaces, \(u^T_tk = p(z^T_tk | u)\) and \(u^T_fk = p(z^T_fk | u)\) are user \(u\)’s topic distributions on the \(k^{th}\) topic.

\(^5\)A hashtag is a word or an unspaced phrase prefixed with the number sign (“#”) in Twitter which shows some meaningful message.
D. Cross-network Topic Association

In this section, we will first describe how we aggregate YouTube video topic distribution to user level and then introduce the association mining scheme between YouTube topic space and Twitter topic space with three kinds of alternative methods. As illustrated in Section III-B and Fig. 3, the semantic-based and network-based associations are both conducted under the same association mining scheme. The only difference lies in the associated topic space in Twitter, i.e., the former one tries to associate with the Twitter user tweet semantic topic space indicated by $U^T_t$, while the latter one aims to associate with the Twitter user followee topic space indicated by $U^T_f$. To focus on the introduction of our cross-network association scheme between two topic spaces, next we will use a unified Twitter topic space indicator $T$ to represent $T_t$ or $T_f$ (e.g., $U^T_T$, $K^T$ and $p(z^T | u)$).

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 $u_i$, we construct the interested video set $V_u \subset V$ from his/her uploaded videos, favorite videos and videos in the playlists. Given YouTube video $v \in V_u$ and its topical distribution $p(z^Y | v)$, through simple derivation, we can calculate user $u_i$’s topical distribution by:

$$p(z^Y_k | u_i) = \frac{N_v(f) + N_v(w)}{N(f) + N(w)} \cdot p(z^Y_k | v)$$

(1)

where $N_v(f)$, $N_v(w)$ denote the total number of keyframes and words in video $v$, $N(f) = \sum_{v \in V_u} N_v(f)$, $N(w) = \sum_{v \in V_u} N_v(w)$ denote the total number of keyframes and words in video set $V_u$. After aggregation, we can obtain the YouTube user topic distribution matrix $U^Y = \{u^Y_1, \cdots, u^Y_{|U^Y|}\}$.

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 $U_o = U^Y \cap U^T$. Viewing as a probabilistic transition problem, the topic association can be calculated by aggregating over all the overlapped users $^6$:

$$a_{ij} = p(z^T_j | z^Y_i) = \sum_{u \in U_o} p(z^T_j | u) \cdot p(u | z^Y_i) = \sum_{u \in U_o} u^T_j \cdot u^Y_i \cdot \frac{p(u)}{p(z^Y_i)}$$

$^6$ The derivation is based on Bayesian rule, which is omitted due to space limitation.
where $p(u)$ is the user prior which we assume having the uniform distribution, the topic prior $p(z^Y_i) = \sum_{u \in \mathcal{U}_o} p(z^Y_i | u) \cdot p(u)$ 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 $A = \{a_{ij}\}_{K^V \times K^V}$. 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 $p(z^Y|u_t)$, his/her Twitter followee topic distribution is estimated as:

$$p(z^T_j | u_t) = \sum_{i=1}^{K^V} a_{ij} \cdot p(z^Y_i | u_t)$$

(2)

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 $U^Y = [U^Y_o, U^Y_{non}]$ and $U^T = [U^T_o, U^T_{non}]$, where $U^Y_o, U^T_o$ 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 $U^Y_o$ to their Twitter distribution $U^T_o$.

Formally, the regression objective function is:

$$\min_A ||U^T_o - AU^Y_o||^2 + \lambda_1 ||A||_q$$

(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 [22]. When $q = 2$, Eqn. (3) is a ridge regression problem with analytical solution as:

$$A = U^T_o U^Y_o (U^Y_o U^Y_o + \lambda_1 I)^{-1}$$

(4)

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

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 $\mathcal{U}^Y_o, \mathcal{U}^T_o$. (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 that 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 $d_Y = \{d_{Y1}, \cdots, d_{YK_Y}\}$ and a Twitter factor $d_T = \{d_{T1}, \cdots, d_{TK_T}\}$ are coupled to the same user attribute $d \in \mathcal{D}$. This can be better understood by analogous to coupled dictionary learning [36]. 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 $D_Y = \{d_{Y1}, \cdots, d_{YjD}\}$, $D_T = \{d_{T1}, \cdots, d_{TjD}\}$ 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:

$$
\min_{D_Y, D_T, S} ||U^Y_o - D^Y S||^2_2 + ||U^T_o - D^T S||^2_2 + \lambda_2 ||S||_1
$$

s.t. $||d_Y||^2_2 \leq 1, ||d_T||^2_2 \leq 1, \forall d \in \mathcal{D}$

where $S = \{s_1, \ldots, s_{|\mathcal{U}_o|}\}$ with $s_i$ be the attribute representation for user $u_i \in \mathcal{U}_o$, the constrain $||d||^2_2 \leq 1$ 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

$$
\min_{D, S} ||\hat{U}_o - \hat{D} S||^2_2 + \lambda_2 ||S||_1
$$

s.t. $||\hat{d}_i||^2_2 \leq 1, \forall i$

where

$$
\hat{U}_o = \begin{bmatrix} U^Y_o \\ U^T_o \end{bmatrix}, \hat{D} = \begin{bmatrix} D^Y \\ D^T \end{bmatrix}
$$

The optimization problem Eqn.(6) can be efficiently solved by the sparse coding algorithm proposed in [22].

**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 $\mathcal{U}^Y_{non}, \mathcal{U}^T_{non}$ also contribute to the user factor discovery in each network, but with no requirement on identical coefficients. Formally,
the optimization objective function is:

$$\min_{D^Y, D^T, S^Y, ST} \||U^Y - D^Y S^Y||_2^2 + ||U^T - D^T S^T||_2^2 + \lambda_3 ||S_o||_1$$

$$+ \lambda_4 ||S_{non}^Y||_1 + \lambda_5 ||S_{non}^T||_1$$

subject to

$$||d^Y||_2^2 \leq 1, ||d^T||_2^2 \leq 1, \forall d \in D$$

(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:

$$\min_{S_o} \||U_o^Y - D^Y S_o||_2^2 + ||U_o^T - D^T S_o||_2^2 + \lambda_3 ||S_o||_1$$

(8)

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

B. Divided factor distribution learning:

$$\min_{S^Y_{non}} \||U_{non}^Y - D^Y S^Y_{non}||_2^2 + \lambda_4 ||S^Y_{non}||_1$$

$$\min_{S^T_{non}} \||U_{non}^T - D^T S^T_{non}||_2^2 + \lambda_5 ||S^T_{non}||_1$$

(9)

This is a multi-task lasso problem and can be solved by the feature-sign search algorithm [22].

C. Coupled user factor update:

$$\min_{D^Y, D^T} \||U^Y - D^Y S^Y||_2^2 + ||U^T - D^T S^T||_2^2$$

subject to

$$||d^Y||_2^2 \leq 1, ||d^T||_2^2 \leq 1, \forall d \in D$$

(10)

This is a quadratically constrained quadratic program problem (QCQP). We utilize an alternative update strategy for solution [37].

The flow of the learning algorithm for LA_all is summarized in Algorithm 1. With the derived user factors $D^Y$ and $D^T$, given a new YouTube user topic distribution $u^Y \in \mathbb{R}^{K_Y \times 1}$, we can estimate the YouTube user factor distribution as:

$$s^* = \min_s ||u^Y - D^Y s||_2^2 + \lambda ||s||_1$$

(11)


**Algorithm 1** Alternative Learning Algorithm for LA_all

**Input:** Users’ topic distributions $U^Y = [U^Y_o, U^Y_{non}], U^T = [U^T_o, U^T_{non}]$; the weighting parameters $\lambda_3, \lambda_4, \lambda_5$.

**Initialize:** $D^Y, D^T$ with the output of Eqn. (5).

**repeat**

1. fix other variables, learn $S_o$ in Eqn. (8).
2. fix other variables, learn $S^Y_{non}, S^T_{non}$ in Eqn. (9).
3. fix other variables, update $D^Y$ and $D^T$ in Eqn. (10).

**until** convergence or reach maximum iteration.

**return:** user factors $D^Y$ and $D^T$.

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

$$u^T = D^T s^*.$$ 

**E. Referrer Identification**

After conducting cross-network topic association in Stage 2, both semantic-based and network-based associations $\mathcal{F}_s, \mathcal{F}_n$ are established, i.e., given a new YouTube user topic distribution $u^Y = p(z^Y|u) \in \mathbb{R}^{K_Y \times 1}$, we can separately reconstruct his/her Twitter tweet semantic topical distribution $u^{T_t} = p(z^{T_t}|u) \in \mathbb{R}^{K_{T_t} \times 1}$ and Twitter followee topical distribution $u^{T_f} = p(z^{T_f}|u) \in \mathbb{R}^{K_{T_f} \times 1}$. In our video promotion problem, given a test YouTube video $v_t$, we simulate a virtual user with identical topic distribution $v^Y_t = p(z^Y|v_t)$ to represent the typical YouTube users liking the video. After distribution transfer, the virtual user’s Twitter tweet semantic topic distribution $v^{T_t}_t = p(z^{T_t}|v_t)$ actually indicates the typical Twitter semantic interest of the video fans, while his/her followee topic distribution $v^{T_f}_t = p(z^{T_f}|v_t)$ actually reflects the video fans’ most probable Twitter following patterns.

On the Twitter side, we construct a popular Twitter followee set $U^{\text{followee}}_t \subset U^{\text{followee}}$ serving as the candidate YouTube video promotion referrers. For each popular followee $u \in U^{\text{followee}}_t$, his/her Twitter topic distributions $u^{T_t}, u^{T_f}$ in both topic spaces are calculated, respectively.

---

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.
For the $u^Tf$ in network-based topic space, we can directly derive his/her topic distribution as:

$$p(z_k^Tf|u) \propto p(u|z_k^Tf) \cdot p(z_k^Tf)$$

where $p(u|z_k^Tf)$ is the topic-word distribution obtained from the network-based twitter user topic modeling in Stage 1, $p(z_k^Tf)$ is the topic prior and can be calculated by aggregating over users. Here $p(z_k^Tf|u)$ actually reflects followee $u$'s popularity in the $k^{th}$ topic; For the $u^Tf$ in semantic-based topic space, since in our video promotion problem the “properness” of Twitter followee is decided by his/her followers, we randomly sample up to 500 followers for each popular followee $u \in U^f_{followee}$ and aggregate all their generated tweets to construct the user document for this Twitter followee. Then the same LDA model as in semantic-based twitter user topic modeling in Stage 1 is applied to this Twitter followee document to obtain his/her topic distribution $u^Tf = p(z^Tf|u)$. Here, $p(z^Tf|u)$ actually indicates followee $u$’s semantic interest upon his/her followers.

**Direct product-based matching** For each topic space separately, 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 $u \in U^f_{followee}$ to promote YouTube video $v_t$ in each topic space is calculated as:

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

$$\text{properness}(u,v_t|T_f) = <v^T_f,u^T_f> = \sum_{k=1}^{K^T_f} v^T_{t,k} \cdot u^T_k$$

(12)

With the properness scores in both topic spaces considered together, the direct product-based matching score can be directly obtained as:

$$\text{properness}(u,v_t) = \text{properness}(u,v_t|T_t) + \text{properness}(u,v_t|T_f)$$

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

**Weighted product-based matching** Notice that each topic may contribute differently to the final matching score between the target video and candidate Twitter followee, especially for the topics distributed in different topic spaces. 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, different topic spaces can be seen as multiple views of the data representation, the referrer identification can be
treated as a retrieval problem. In light of this, we design a training scheme and adopt ranking SVM [17] for topic selection.

Ranking SVM model is with the form as:

$$g(\cdot, \cdot) = h \cdot \phi(\cdot, \cdot)$$  \hfill (13)

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

In our problem, we define the feature mapping function for each video-followee pair \( u, v \) as the concatenation of the vector products between video and followee distributions in both topic spaces:

$$\phi(u, v) = \{ u^{T_i} \odot v^{T_i}, u^{T_j} \odot v^{T_j} \}$$

where,

$$u^{T_i} \odot v^{T_i} = \{ u_1^{T_i} \cdot v_1^{T_i}, ..., u_K^{T_i} \cdot v_K^{T_i} \}$$

$$u^{T_j} \odot v^{T_j} = \{ u_1^{T_j} \cdot v_1^{T_j}, ..., u_K^{T_j} \cdot v_K^{T_j} \}$$

Among the above equations, \( \odot \) indicates the element-wise multiplication. To obtain the actual 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:

$$\text{precision}(v, u) = \frac{|U_v \cap U_u^{\text{followee}}|}{|U_u^{\text{followee}}|}$$

$$\text{recall}(v, u) = \frac{|U_v \cap U_u^{\text{followee}}|}{|U_v|}$$

$$\text{GT-properness}(v, u) = \frac{2}{\text{precision}(v, u)^{-1} + \text{recall}(v, u)^{-1}}$$  \hfill (14)

where \( U_v \) is the set of users showing interest in \( v \), \( U_u^{\text{followee}} \) is the follower set of \( u \). We can see that recall actually concerns with coverage of the interested YouTube audiences, while precision is in charge

8 In this stage, we mainly aim to demonstrate the superiority of integrating information from different topic spaces. But we don’t elaborate on different training schemes for combining them.

9 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 \( U_v \) and \( U_u^{\text{followee}} \) are counted based on the known overlapped user set \( U_o \).
Fig. 6. The perplexities for different topic numbers on YouTube and Twitter: (a) On YouTube video multimodal topic space; (b) On Twitter semantic-based topic space; (c) On Twitter network-based topic space.

TABLE II
STATISTICS OF OUR DATASET.

| $|\mathcal{U}_y|$ | $|\mathcal{U}_t|$ | $|\mathcal{U}_f|$ | $|\mathcal{V}|$ | $|\mathcal{E}_{tweet}|$ | Avg. $|\mathcal{U}_f^{followee}|$ |
|----------------|----------------|----------------|-------------|----------------|----------------|
| 38,540         | 39,400         | 11,850         | 2,280,129   | 31,818,609     | 891.1          |

of the virtual cost. With the learnt $h^*$, the properness of Twitter followee $u \in \mathcal{U}_f^{followee}$ for test video $v_t$ is calculated as:

$$properness(u, v_t) = h^* \cdot \phi(v_t^T, u^T_f; v_t^T, u^T_f).$$

IV. EXPERIMENTS

A. 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. Therefore, we started from Google+ website and randomly selected some seed users, then the snowball sampling method is utilized to collect 143,259 Google+ users from their

TABLE III
VISUALIZATION OF DISCOVERED TWITTER TWEET SEMANTIC TOPICS IN TWITTER SEMANTIC-BASED TOPIC SPACE.

| Topic | The top-5 probable tweet words in terms of $p(w|z_{t}^T)$ |
|-------|---------------------------------------------------------|
| #12   | people news government vote state                      |
| #57   | google android apple phone windows                     |
| #3    | game team cup win WorldCup                             |
### TABLE IV

**Visualization of discovered YouTube video topics.**

<table>
<thead>
<tr>
<th>Topic #1</th>
<th>Word</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>gameplay, xbox, playstation, gaming, minecraft</td>
<td>“Epic Mods - MW2 MOD IN CoD4”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“HEXXIT COOP ep7 w/ Double”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Halo 4 Drift Multiplayer Map”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Topic #17</th>
<th>Word</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>history, german, berlin, germany, poetry</td>
<td>“GEH STERBEN, DU OFFER!!!”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Syrien - Wahrheit ber das Massaker”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Volker Pispers - Einzeltater”</td>
</tr>
</tbody>
</table>

Social connections. The external account links of these users are kept, among which 38,540 provide YouTube account, 39,400 provide Twitter account, 11,850 provide both accounts. 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, generated tweets information and user profiles via Twitter API. Table II summarizes the key statistics.

---

10 User linkage mining is a separate topic in cross-network analysis [25]. In our work, to guarantee a promising overlapped user resource, we leverage user self-provided account links on Google+.

11 The most recent 1,000 tweets are downloaded for each Twitter user.

12 Avg. $|\mathcal{U}_\text{followee}|$ is the average number of followees over all the examined Twitter users, $|\mathcal{U}_\text{tweet}|$ is the total number of tweets downloaded.
B. Heterogeneous Topic Modeling

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 unobserved data [4]. More formally, for a test set of $M$ documents, the perplexity is:

$$\text{perplexity}(D_{test}) = \exp\left\{-\frac{1}{M} \sum_{d=1}^{M} \frac{\log p(w_d)}{N_d}\right\}$$

where $N_d$ is the total number of words in document $d$, $p(w_d)$ is the probability of the unseen document $d$. The lower the perplexity score is, the better the performance. We test the perplexity with different topic number $K^Y$ on 490,000 held-out YouTube videos, $K^{T_t}$ and $K^{T_f}$ on 9,400 held-out Twitter users, respectively. The perplexity scores on different topic numbers are shown in Fig. 6. 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 $K^{T_t} = 60$, $K^{T_f} = 80$ for Twitter.

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 IV 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 IV. 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 III shows three sampled Twitter tweet semantic topics. For each topic, the top-5 probable tweet words are provided. It is also easy to interpret the domain knowledge associated with each topic and some topics may capture similar textual semantics as in YouTube video topics (e.g., #2 for game-related concept). Table V 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

---

13 Hyperparameters are fixed as $\alpha = 0.8$ and $\beta = 0.1$ according to the empirical expectation for the output distribution [14].
TABLE V

VISUALIZATION OF DISCOVERED Twitter FOLLOWEE TOPICS IN Twitter NETWORK-BASED TOPIC SPACE.

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

TABLE VI

COVER RATIO OF THE POPULAR OFFICIAL Twitter ACCOUNTS.

<table>
<thead>
<tr>
<th>Top-k</th>
<th>Top-10</th>
<th>Top-20</th>
<th>Top-30</th>
<th>Top-40</th>
<th>Top-50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cover ratio</td>
<td>100%</td>
<td>80%</td>
<td>73.3%</td>
<td>65%</td>
<td>62%</td>
</tr>
</tbody>
</table>

topic #10 consisting of Forbes influencers, and even the geographic topic #38 with the top followees all coming from Berlin. Moreover, we also aggregate the top-50 Twitter followees in each Twitter followee topic and examine how much of these discovered users cover the most popular Twitter accounts listed in official site “http://twitaholic.com/”. Table VI shows the result and we can see that the obtained users can cover most of the actual top Twitter followees, which indicates the wide domain expertise of the derived Twitter followee topic space. Twitter users’ joint following patterns are well captured in modeling the follower-followee relationship, which is very important to the subsequent promotion application.

C. Cross Network Topic Association

1) Experimental Setting: Given a user with his/her YouTube topic distribution $u^Y$, the goal of Stage 2 is to estimate his/her Twitter topic distribution $u^T$ ($u^{T_1}$ or $u^{T_2}$ for different kinds of association). Therefore, we utilize Mean Absolute Error (MAE) as the evaluation metric. MAE is a statistical accuracy metric which is widely used in predicting accurate error. We randomly select half of the overlapped users to construct the test set $U_{test}$, and the rest overlapped users and non-overlapped users as the training
set. The semantic-based association and network-based association are evaluated under the MAE metric, separately. Specially, MAE is calculated over all topics of each test user as:

$$\text{MAE} = \frac{\sum_{u \in \mathcal{U}_{\text{test}}} \sum_{k=1}^{K_T} |\hat{u}_k^T - u_k^T|}{|\mathcal{U}_{\text{test}}| K_T}$$

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 $\lambda_1$ in Eqn. (3) by grid search and 5-fold cross validation. Tuning parameters $\lambda_2$ 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 (1) for semantic-based association: $\lambda_1 = 0.1, \lambda_2 = 0.01, \lambda_3 = 0.005, \lambda_4 = 0.01, \lambda_5 = 0.01$; (2) for network-based association: $\lambda_1 = 0.1, \lambda_2 = 0.2, \lambda_3 = 0.1, \lambda_4 = 0.01, \lambda_5 = 0.01$. 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) [15], are widely used to evaluate how well the observed data can be reconstructed from the learnt dictionary. Therefore, for each kind of cross-network association, we conduct K-means on YouTube and Twitter user distributions $U_Y, U_T$ with the identical cluster number $|\mathcal{D}|$. In Fig. 7 we draw the curve of WCSS sum on the two networks w.r.t. the change of $|\mathcal{D}|$ for the semantic-based and network-based associations, respectively. We choose $|\mathcal{D}| = 500$ in semantic-based association and $|\mathcal{D}| = 300$ in network-based association when the aggregated reconstruction error decreases to a steadily low level.

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. (1) For network-based association, 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 IV and V. 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. (2) For semantic-based association, among the $K_Y \times K_T = 2,400$ topic association pairs, the most significant are $\{z_{2}^Y, z_{57}^T\}$ and $\{z_{30}^Y, z_{12}^T\}$, where the first one is about the semantic topic of mobile phone and the second one is related with the semantic topic of American president election. Here we only visualize the involved Twitter
tweet semantic topics in Table III for space consideration. With both semantic-based and network-based associations considered, we can capture a wide coverage of correlation between YouTube and Twitter from different aspects. 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 for each kind of association is shown in Fig. 8. Several observations are made. (1) With much lower MAE for network-based association than semantic-based association, it may prove to be a more appropriate way to associate Twitter space by leveraging its follower-followee network structure. This also again validates the flexibility of the network-based association. (2) For the regression-based methods, the L1 method can obtain better performance than the L2 method. This may be ascribed to the fact that not every YouTube topic has an associated Twitter topic and some topic pairs between YouTube and Twitter spaces are just irrelevant. By introducing the sparse L1 regularization to the association matrix, the L1 method can well weaken the influence of the typical outliers. (3) The latent attribute-based methods (LA_overlap, LA_all) outperform the explicit
association matrix-oriented methods \((TP, \text{Regression})\) for both kinds of association. 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. (4) 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.

\[\text{D. Twitter Referrer Identification}\]

In this section, we first introduce the experimental setting and how we evaluate the results of our final problem. Then the algorithm performance and comparison for different possible methods are given. To further explore how different kinds of cross-network association in Stage 2 contribute to the algorithm performance, we finally evaluate the results for each kind of association separately.

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 \(V_t\). Meanwhile, 21,276 Twitter followees who are followed by more than 50 users construct the candidate Twitter followee set \(U_{follower}^t\).

We use Normalized Discounted Cumulative Gain \((NDCG)\) and user coverage \((Cov)\) as the evaluation metrics, which are defined as,

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

\[
Cov@k = \frac{\left| \bigcup_{j=1}^{k} U_{follower}^j \cap U_v \right|}{|U_v|}
\]

where \(NDCG\) is a ranking metric which is widely used in retrieval problems \([16]\), \(rel(\cdot)\) 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 \(rel(\cdot)\). In this way, the \(NDCG\) metric actually considers both the precision and recall of the video promotion problem. \(Z\) is a normalization factor which can be obtained by calculating the ideal \(NDCG\) based on the known overlapped user set \(U_o\). For the user coverage metric, \(U_{follower}^j\) is the follower set of the \(j\)th identified Twitter followee while \(U_v\) is the ground-truth set of users showing interest in target video \(v\). Indeed, the user coverage metric \((Cov)\) reflects the actual coverage of the interested video audience following the top-ranked identified Twitter followees. In other words, the \(Cov\) metric mainly examines the total recall of the video promotion problem when recommending a subset of \(k\) identified Twitter followees.
2) Evaluation for Different Methods: To validate the effectiveness of our cross-network association solution, we compare our approach with different possible methods as follows,

- **Random**: randomly select $k$ followees from $U_t^{\text{followee}}$;
- **Popularity**: select $k$ popular Twitter followees with the most #followers;
- **Semantic Match**: directly textual matching between Twitter followee’s tweets and YouTube videos textual metadata. Vector Space Model (VSM) is used to represent the Twitter followees and target YouTube video and we select the top-$k$ followees by sorting the cosine similarity between the TF-IDF coded vectors;
- **Semantic Association**: the same as our framework, but only the semantic-based association is conducted by associating Twitter user tweet topic space (for various alternative methods in Stage 2 and Stage 3, we only show the result with the best performance);
- **Network Association**: the same as our framework, but only the network-based association is conducted by associating Twitter followee topic space (also only show the best performance);
- **Both Associations+Direct**: Both kinds of association are conducted in Stage 2 and direct product-based matching is utilized in Stage 3;
- **Both Associations+Weighted**: Both kinds of association are conducted in Stage 2 and weighted product-based matching is utilized in Stage 3.

We show $\text{NDCG}@5$ and $\text{Cov}@5$ for different methods in Fig. 9 and Fig. 10, respectively. It is observed that: (1) By identifying popular Twitter followees with the most #followers, Popularity is able to achieve a much higher user coverage ($\text{Cov}$) than all the other methods. While high #follower guarantees the coverage of potential viewers (recall), the retrieved follower set is expected to also include many uninterested users (precision), which results in a relatively low $\text{NDCG}$ value. It is also rather difficult to get these popular
followees to help due to the high cost in practical situation. (2) *Semantic Match* fails to identify the optimal Twitter referrer. Directly matching from semantic level cannot guarantee the popularity of the identified Twitter followees, which may result in the low user coverage (Cov) value. The discrepancy in topic granularity makes it impractical to match Twitter followees’ interest to YouTube videos only from the semantic level and the noise or outlier in user/video textual metadata may largely influence the matching process. (3) *Network Association* achieves better performance than *Semantic Association* on both NDCG@5 and Cov@5 metrics. This again validates the flexibility of the network-based association, which can better handle the topic granularity discrepancy issue between different social platforms. (4) By considering both the semantic-based and network-based associations together, *Both Associations+Direct* and *Both Associations+Weighted* outperform *Semantic Association* and *Network Association* in which only one kind of association is investigated. (5) *Both Associations+Weighted* achieves better performance than *Both Associations+Direct*. This demonstrates the advantage of topic weight optimization. This method also obtains more superior results on both evaluation metrics compared with other methods.

3) *Further Exploration for Each Kind of Association Separately*: To investigate how different kinds of cross-network association and alternative methods will influence the final performance of our video promotion problem, we further conduct a comparison experiment between *Semantic Association* and Net-
work Association for different alternative methods in our solution. The following settings are considered for comparison:

- **Regression+Direct**: Association mining by Regression_1, matching by Direct product;
- **Regression+Weighted**: Association mining by Regression_1, matching by Weighted product;
- **LA_all+Direct**: Association mining by LA_all, matching by Direct product;
- **LA_all+Weighted**: Association mining by LA_all, matching by Weighted product.

The results are shown in Fig. 11 and Fig. 12, respectively. The combinations of different alternative methods in Stage 2 and Stage 3 for each kind of association are evaluated. We can see that: (1) For all the same combinations, Network Association consistently outperform Semantic Association, which demonstrates the effectiveness of the network-based association. Possible reasons are that network structure is more stable as a user interest indicator than noisy tweet activity in Twitter and the consideration of user’s network structure is significant in our propagation scenario. (2) The consideration of topic weight optimization in Weighted product settings can always improve the performance over the Direct product settings. Different topics and topic spaces should contribute differently in view of referrer identification and this can be learnt under a supervised ranking scheme in Weighted product settings. (3) In terms of $NDCG$ metric, conducting association mining 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 association mining contributes to improved referrer identification. However, this superiority does not apply to the $Cov$ metric. This may be due to the reason that the follower overlap issue for a group of Twitter followees is not directly considered in our framework, which is critical in $Cov$ metric.
V. Conclusions

We have proposed an overlapped user-based association solution framework, to address the novel cross-network YouTube video promotion problem. To better capture the cross-network association from different perspectives, we conduct both semantic-based and network-based associations in a unified ranking scheme. 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.

REFERENCES


**Ming Yan** received the B.E. degree from Huazhong University of Science and Technology, Wuhan, China in 2011 and is currently pursuing the Ph.D. degree at the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China. His research interest lies in user-centric social media mining and application, knowledge mining and cross-network analysis.

**Jitao Sang** is an assistant professor in National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences (CASIA). He received the B.E. degree from SouthEast University in 2007, and the Ph.D. degree from CASIA in 2012. He was awarded the Special Prize of President Scholarship by Chinese Academy of Sciences. His research interests include multimedia content analysis, social media mining and social network analysis. Dr. Sang has published several refereed research papers in these areas and co-authored the Best Student Paper in Internet Multimedia Modeling 2013 and the Best Paper Finalist papers in ACM Multimedia 2012, 2013. He has served as guest editors and organizing committee members in several journals and conferences.
Changsheng Xu (M 97 - SM 99 - F 14) is a Professor with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China, and an Executive Director of the China-Singapore Institute of Digital Media, Singapore. His current research interests include multimedia content analysis/indexing/retrieval, pattern recognition, and computer vision. He holds 30 granted/pending patents and has published over 200 refereed research papers. Dr. Xu is an Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA and ACM Transactions on Multimedia Computing, Communications. He served as a Program Chair of ACM Multimedia in 2009. He has served as an Associate Editor, Guest Editor, General Chair, Program Chair, Area/Track Chair, Special Session Organizer, Session Chair and TPC Member for over 20 prestigious IEEE and ACM multimedia journals, conferences, and workshops. He is an ACM Distinguished Scientist.

M. Shamim Hossain (S03CM07CSM09) received the Ph.D. degree in electrical and computer engineering from the University of Ottawa, Ottawa, ON, Canada.

He is an Associate Professor with King Saud University, Riyadh, Saudi Arabia. He has authored or co-authored more than 70 publications including refereed IEEE/ACM/Springer/Elsevier journals, conference papers, books, and book chapters. His research interests include serious games, cloud and multimedia for health care, resource provisioning for big data processing on media clouds, and biologically inspired approach for multimedia and software system.

Dr. Hossain has served as a member of the organizing and technical committees of several international conferences and workshops. He served as a Co-Chair of the 1st, 2nd, 3rd, 4th, and 5th IEEE ICME Workshop on Multimedia Services and Tools for E-Health. He served as a Co-Chair of the 1st Cloud-Based Multimedia Services and Tools for E-Health Workshop 2012 with ACM Multimedia. He currently serves as a Co-Chair of the 4th IEEE ICME Workshop on Multimedia Services and Tools for E-Health. He is on the editorial board of the Springer International Journal of Multimedia Tools and Applications. He was on the editorial board of some journals including the International Journal of Advanced Media and Communication. Previously, he served as a Guest Editor of the IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE and Springer Multimedia Tools and Applications. Currently, he serves as a Lead Guest Editor of the Elsevier Future Generation Computer Systems, International Journal of Distributed Sensor Networks, and Springer Cluster Computing. He is a member of ACM and ACM SIGMM.