Hierarchical Community-Level Information Diffusion Modeling in Social Networks

Yuan Zhang
Department of Machine Intelligence
Key Laboratory of Machine Perception (MOE)
Peking University, China
yuan.z@pku.edu.cn

Tianshu Lyu
Department of Machine Intelligence
Key Laboratory of Machine Perception (MOE)
Peking University, China
lyutianshu@pku.edu.cn

Yan Zhang
Department of Machine Intelligence
Key Laboratory of Machine Perception (MOE)
Peking University, China
zhy@cis.pku.edu.cn

ABSTRACT

Recently, online social networks are becoming increasingly popular platforms for social interactions. Understanding how information propagates in such networks is important for personalization and recommendation in social search.

In this paper, we propose a Hierarchical Community-level Information Diffusion (HCID) model to capture the information diffusion process in social networks. We introduce the notion of users' *topic popularity* as to enable our model to depict the information diffusion process which is both topic-aware (which topic the information is concerned with) and source-aware (where the information comes from). Instead of assuming homogeneity of social communities, we propose the notion of *community hierarchy*, where information diffusion across inter-level communities is uni-directional from the higher levels to the lower ones.

We design a Gibbs sampling algorithm to infer model parameters and propose prediction methods for two information diffusion prediction tasks, the retweet prediction and the cascade prediction. Comparison experiments are conducted on two real datasets. Results show that our model achieves substantial improvement compared with the existing work.

CCS CONCEPTS

•Applied computing →Sociology; •Human-centered computing →Social networking sites; •Information systems →*Data mining;*

KEYWORDS

social networks; communities; information diffusion

ACM Reference format:
Yuan Zhang, Tianshu Lyu, and Yan Zhang. 2017. Hierarchical Community-Level Information Diffusion Modeling in Social Networks. In *Proceedings of SIGIR '17, August 7–11, 2017, Shinjuku, Tokyo, Japan, ,* 10 pages.
DOI: http://dx.doi.org/10.1145/3077136.3080784

SIGIR '17, August 7–11, 2017, Shinjuku, Tokyo, Japan
© 2017 ACM. 978-1-4503-5022-8/17/08…$15.00
DOI: http://dx.doi.org/10.1145/3077136.3080784

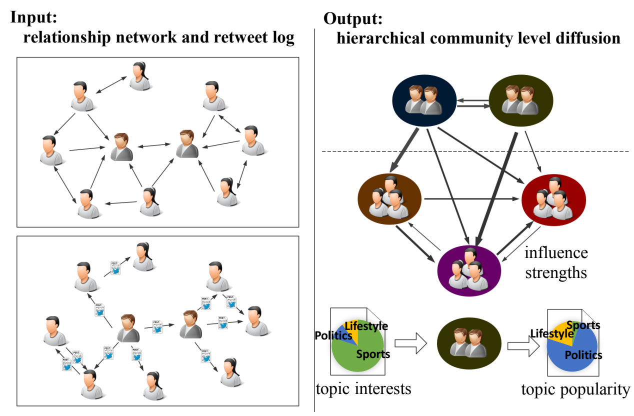

Figure 1: Hierarchical Community Level Diffusion.

1 INTRODUCTION

In the era of Web 2.0, online social networks (OSN), such as Twitter, Facebook and Weibo, are becoming increasingly popular platforms for social interactions and communication. Social network users form a large-scale complex network by establishing who-follows-who relationships/friendships, and generate textual contents by tweeting/posting. They also share semantic information through their social network connections, e.g. retweeting/reposting, which accounts for the majority of information diffusion in social networks. In this paper, we focus on modeling such information diffusion processes to gain better insight on the information spreading mechanisms along with users' topical and social preferences, which are useful in many IR tasks such as *personalized social search*.

Related studies can be roughly divided into two main categories: one is information diffusion modeling [3, 17, 19, 23, 36, 37], the other is semantic analysis combining textual content with network structures [4, 7–9, 16, 22, 24]. Although much progress has been made, the existing work still suffers from several limitations:

- Most work in literature models information diffusion at the individual level. However, the high volatility of user behaviors renders it difficult to accurately uncover diffusion patterns for individual level models [19]. Moreover, those models, especially influence-based ones e.g. [23, 34, 37], become computationally prohibitive to inference when dealing with large scale networks usually with more than millions of nodes and billions of links.

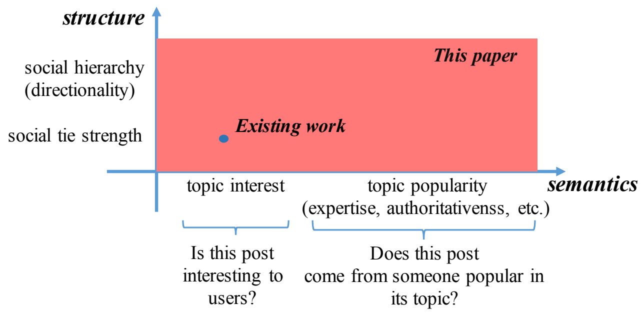

Figure 2: Two-dimension interleaving space for information diffusion phenomena. Most existing work only tackles one dimension or one point in the space.

- Few models holistically depict the retweeting behaviors, which explicitly indicate information diffusion from source users to target users (i.e. retweeters). For example, Hu et. al. [19] model the retweet network and posts separately, ignoring semantic information on retweet links as well as posts' retweeting source, which fails to capture the whole retweeting process and has weaker performance as showed in Section 3.2. [35, 37] have studied retweeting behaviors systematically; nonetheless, the discriminative methods they use make it only applicable to specific tasks, such as the retweet prediction, but not to the others.

To overcome those limitations, we propose a unified information diffusion approach, HCID (Hierarchical Community-level Information Diffusion), to jointly model information diffusion processes in two dimensions, *network structure* and *information semantics* (see Figure 2). The proposed approach models information diffusion at the community level similarly to [19], yet our contribution extends the notion of communities to better capture information diffusion in those two dimensions respectively.[1]

First, the traditional concept of communities is typically only concerned with social tie strengths to capture locally dense subgraph structures. However, social science studies reveal that social communication exhibits hierarchy, in the sense that people from higher level communities barely receive information from lower level communities. For instance, influential users (celebrities, amusing story publishers etc.) are more popular and hence have more followers, but they seldom follow "ordinary people" and cautiously retweet their posts due to high social status and concerns about public opinions in twitter-like microblogs. We plot the numbers of users' followers (i.e., in-degrees, which reflect users' social status to a certain extent) in unilateral and reciprocal follow relationships in Weibo[2] in Figure 3. We can find that one tends to follow other users with higher out-degrees (which is also studied as *disassortative degree mixing* [29]), while reciprocal relationships are more likely to establish between users with similar numbers of followers (which is called *assortativeness* [25] in network science).

[1]The father of *mass communications*, Wilbur Schramm, argues that it is no accident that "community" and "communication" have the same word roots [28].
[2]Sina Weibo is a popular microblogging service in China.

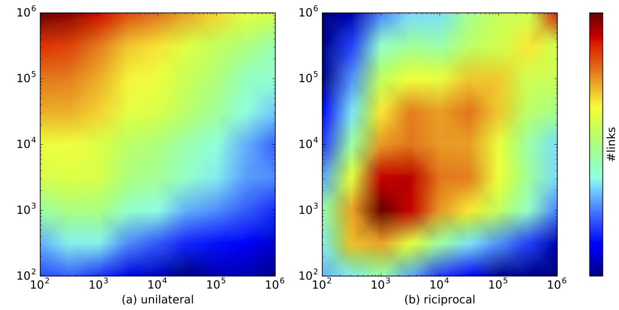

Figure 3: Follow relationship hierarchy in Weibo. The horizontal axis and the vertical axis denote the numbers of followers for users on the outward and inward sides of follow relationships, respectively.

Thus, we introduce the notion of *community hierarchy*, which defines the direction of information flows at the community level and enables our model to capture information diffusion patterns of higher complexity. Further, empirical study verifies this intuition and shows substantial model performance improvement on various evaluation metrics.

Second, social communities are associated with similar topical preferences in our model. Most approaches, concerning users' topic interests e.g. [16, 19, 22], assume that every user has a single topic distribution, serving as both his/her topic interests and popularity. However, a user's interest in a topic (as a receiver) is not necessarily the same as others' interests towards him/her (as a sender) in that topic. For example, a famous politician in Twitter might take strong interests in American football, thereby retweeting a large amount of posts[3] about NFL games and only a few pieces of politics related posts. Most people usually are not interested in the NFL news he shares but those political opinions that account for only a small proportion of his total posts. Therefore, we propose the notion of communities' *topic popularity* to capture this kind of source-aware information diffusion phenomenon upon different topics.

Figure 1 summaries our proposed model. The input is a relationship network along with retweeting logs, each of which contains the textual content, the source user and the target user (retweeter) of a retweet. Our goal is to understand the mechanism of information diffusion across communities by extracting hierarchical communities with their interests and popularity in different topics.

To conclude, the major contributions of this paper are:

- Our model extends the concept of social communities to better model information diffusion processes. We take community hierarchy into consideration in addition to social tie strengths. Further, we incorporate communities' topic popularity along with their topic interests into our model to capture the topic-aware and source-aware information diffusion phenomenon.
- We design a constrained Gibbs sampling algorithm to estimate model parameters, which is computationally linear w.r.t. the input network and retweeting logs. Those interpretable parameters can be exploited in several application tasks.

[3]We use the words "post" and "tweet" interchangeably in this paper.

Table 1: Notations used in this paper.

Symbols	Descriptions
U, V, C, K	number of users, vocabularies, communities, and topics
E_u, D_u	number of retweet logs of user u as a retweeter, and links from user u
$\boldsymbol{\pi_u}$	*community membership*, multinomial distribution over communities of user u
$\boldsymbol{\theta_c}$	*topic interests*, multinomial distribution over topics of community c
$\boldsymbol{\theta'_c}$	*topic popularity*, multinomial distribution over topics of community c
$\boldsymbol{\psi_k}$	multinomial distribution over words specific to topic k
$\boldsymbol{\phi}$	multinomial distribution over topics of the total posts
$\eta_{cc'}$	topic-irrelevant social influence to community c from community c'
I_e	indicator of the existence of link e
I_d	indicator of the existence of retweet d
s_e, s'_e	communities associated with follower and followee of link e
c_d, c'_d	communities associated with target user and source user of retweet d
z_d	topic of retweet d
w_{dl}	the l-th word of retweet d
λ_0, λ_1	Beta priors on $\boldsymbol{\eta}$
$\rho, \alpha, \alpha', \gamma, \beta$	Dirichlet priors on $\boldsymbol{\pi_u}, \boldsymbol{\theta_c}, \boldsymbol{\theta'_c}, \boldsymbol{\psi_k}, \boldsymbol{\phi}$

- Through experiments on real datasets, we demonstrate substantial predictive performance improvement achieved by our model compared with the existing work, and interesting insights concerning community hierarchy in microblog social networks.

The rest of this paper is organized as follows. Section 2 introduces our proposed model and its major application in information diffusion prediction, while Section 3 reports our experimental analysis. We review literature in Section 4 and conclude our work in Section 5.

2 MODEL

2.1 Formulation

We use $G = (\mathcal{U}, \mathcal{E})$ to denote the relationship network, where $\mathcal{U}$ is the set of all users, $\mathcal{E}$ is the set of directed social links (e.g. who-follows-who relationships and retweet networks). $\mathcal{D}$ is used to represent the set of retweet logs. Each retweet log $d = (u, v, \mathbf{w_d}) \in \mathcal{D}$ denotes a post retweeted by the *target user* u from the *source user* v with textual content $\mathbf{w_d}$.

Definition 2.1. **Community.** A social network consists of C communities denoted as $c \in [1, 2, ..., C]$, where community members intensively interact with each other and share similar social preferences. Each user $u \in \mathcal{U}$ may take part in multiple communities in different contexts, according to a multinomial distribution $\boldsymbol{\pi_u}$,

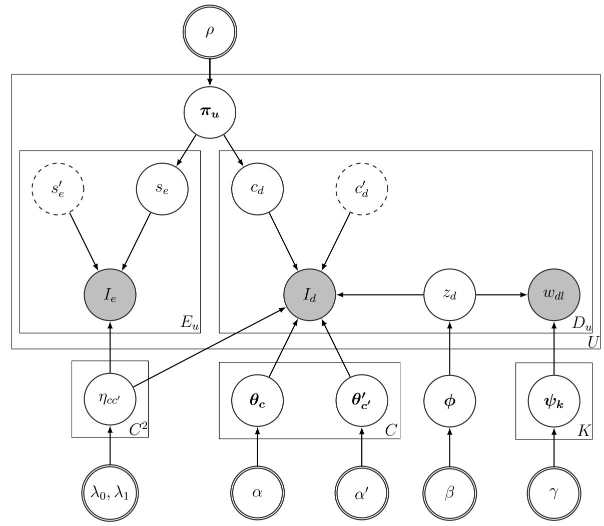

Figure 4: The Graphical Model Representation.

where each component π_{uc} denotes user u's tendency toward the membership of community c.

Definition 2.2. **Topic.** A topic $k \in [1, 2, ..., K]$ is represented by a multinomial distribution over the vocabulary, denoted as $\boldsymbol{\psi_k}$. Every post $d \in \mathcal{D}$ is associated with a topic z_d according to its textual content $\mathbf{w_d}$.

Definition 2.3. **Topic Interest and Topic Popularity.** Every community c has two components: a multinomial distribution $\boldsymbol{\theta_c}$, where θ_{ck} represents the *topic interest* of community c toward topic k; and a multinomial distribution $\boldsymbol{\theta'_c}$, where θ'_{ck} represents the *topic popularity* of community c with regard to topic k.

Definition 2.4. **Community-level Influence Strength.** There is a Bernoulli distribution $\eta_{cc'}$ between any two communities c and c', representing their topic-irrelevant *influence strength*.

Definition 2.5. **Community Hierarchy.** We divide communities into L hierarchical levels, denoted as $l \in [1, 2, ..., L]$, where *level 1* representing the lowest level and *level L* the highest. Each community c is pre-assigned to a specific hierarchical level l_c.

All the notations used in this paper are listed in Table 1.

2.2 Model Description

Users take part in different communities in different settings. Therefore, we employ the mixed-membership approach [1]: each user u is associated with a multinomial membership distribution vector π_u, where each component $\pi_{uc} = P(c|u)$ denotes the probability that user u acts as a member of community c.

We also incorporate probabilistic semantic analysis into our model. Each retweet log $d = (u, v, \mathbf{w_d}) \in \mathcal{D}$ contains a bag of words $\mathbf{w_d}$. We assign a topic z_d to each d. The bag of words $\mathbf{w_d}$ is then drawn from the corresponding word distribution $\boldsymbol{\psi_{z_d}}$. Unlike [19, 38], the topics are drawn from a consistent topic distribution $\boldsymbol{\phi}$ instead of a personalized topic distribution (e.g., $\boldsymbol{\phi_c}$), because we focus on retweets instead of original posts in our model and, of course, users who retweet those posts cannot control their textual contents, which are usually determined by combined effect of the entire social network.

In our model, every time a user u establishes a social link e to user v, user u chooses to act as a member of a specific community s_e, and user v is regarded as a member of specific community s'_e. The probability of the existence of link e is determined by the community-level influence strength $\eta_{s_e s'_e}$.

Likewise, each retweet log d is assigned to two communities c_d and c'_d, one for the target user (retweeter) u and one for the source user v, denoting their community membership when taking this retweeting action. In other words, inter-community interactions serve as information diffusion *channels* in our model. In order to capture the topic-aware and source-aware retweeting mechanism, we assume that the retweeting decision depends on three factors: the topic-irrelevant influence strength $\eta_{c_d c'_d}$, the topic interest of the target community θ_{c_d}, and the topic popularity of the source community $\theta'_{c'_d}$. The first factor serves as the potential likelihood for retweet actions to occur, while the latter two determine whether the potential likelihood is activated due to retweeters' interests in both the topic and the information source.

Here, we utilize both social links and retweet interactions to model information diffusion, under the observation that retweet interactions reflect information diffusion more accurately, but are sparser than social links. Informally, social links serve as "priors" for retweet interactions and become trivial when retweet logs are sufficient.

Due to the sparsity of social networks, we only model positive links and existing retweet logs; s_e, s'_e and c_d, c'_d exist if and only if $e \in \mathcal{E}$ and $d \in \mathcal{D}$, respectively. As in [18, 19], we use a $Beta(\lambda_0, \lambda_1)$ prior on each $\eta_{cc'}$ and negative samples are implicitly modeled in the *beta-binomial* hyper-parameters $\lambda_0 = \kappa \cdot \ln(N_{neg}/C^2)$ and $\lambda_1 = 0.1$, where $N_{neg} = U(U-1)(1 + D/E) - E - D$ and κ is a tunable weight. Thus, linear model complexity w.r.t the number of links and retweet logs can be achieved.

Finally, we address the notion of community hierarchy. Every community c is pre-assigned to a hierarchical level l_c. A social link e can only be established when $l_{s_e} \le l_{s'_e}$, and a retweet action d can only occur when $l_{c_d} \le l_{c'_d}$. That is, we assume information flows downwards or horizontally, not upwards, because of social status. Yet, such information diffusion processes of high complexity as loops can still be captured thanks to multiple diffusion channels established between multiple community memberships. For example, a user may receive a piece of information within a lower level community (possibly from a higher one) and diffuse this piece of information as a member of a higher level community she also belongs to. For the same reason, the possibility that "high level" users may retweet some high-quality posts from "low level" ones is not ruled out.

Note that, in our model, we use with the same conjugate Beta priors on $\boldsymbol{\eta}$ to model influence strengths across both inter- and intra- level communities for tractability and simplicity. The posterior influence strengths across two anti-hierarchy communities are minimized in the model inference stage, as explained later in Section 2.3. Also, since community detection is an unsupervised procedure in our model, we can arbitrarily assign hierarchical levels to communities according to a pre-set partition. For instance, if we want to associate 20% of communities with the highest level, we can simply assign *level L* to the first 20% communities (i.e. let $l_c = 1, \forall c \in [1, ..., \lfloor C/5 \rfloor]$), thanks to symmetry.

Figure 4 shows the graphical representation of our proposed model. The generative process of all social links and retweet logs is summarized in Algorithm 1[4].

Algorithm 1 Generative process.

(1) For each community $c = 1, 2, ..., C$,
  (a) Draw topic interest distribution, $\boldsymbol{\theta_c} \sim Dir(\alpha)$.
  (b) Draw popularity distribution, $\boldsymbol{\theta'_c} \sim Dir(\alpha')$.
  (c) For each community $c' = 1, 2, ..., C$,
    (i) Draw the community-level influence strength, $\eta_{cc'} \sim Beta(\lambda_0, \lambda_1)$.
(2) For each topic $k = 1, 2, ..., K$,
  (a) Draw the multinomial distribution over vocabularies, $\boldsymbol{\psi_k} \sim Dir(\gamma)$.
(3) For each user $u = 1, 2, ..., U$,
  (a) Draw the community membership distribution, $\boldsymbol{\pi_u} \sim Dir(\rho)$.
  (b) For each link $e = (u, v) \in \mathcal{E}_u$,
    (i) Draw user u's community, $s_e \sim Cat(\boldsymbol{\pi_u})$.
    (ii) Draw user v's community, $s'_e \sim Cat(\boldsymbol{\pi_v})$.
    (iii) Draw the existence indicator, $I_e \sim Ber(\eta_{s_e, s'_e})$.
  (c) For each retweet log $d = (u, v, \mathbf{w_d}) \in \mathcal{D}_u$,
    (i) Draw user u's community, $c_d \sim Cat(\boldsymbol{\pi_u})$.
    (ii) Draw user v's community, $c'_d \sim Cat(\boldsymbol{\pi_v})$.
    (iii) Draw topic indicator, $z_d \sim Cat(\boldsymbol{\phi})$
    (iv) Draw textual content, $\mathbf{w_d} \sim Mul(\boldsymbol{\psi_{z_d}})$.
    (v) Draw the existence indicator, $I_d \sim Ber(\eta_{c_l, c'_l} \theta_{c_d z_d} \theta'_{c'_d z_d})$.

2.3 Inference and Parameter Estimation

In this section, we show how to estimate the latent space parameters (i.e. $\boldsymbol{\pi}$, $\boldsymbol{\theta}$, $\boldsymbol{\theta'}$, $\boldsymbol{\eta}$, $\boldsymbol{\phi}$, $\boldsymbol{\psi}$) from observed data (i.e. $G(\mathcal{U}, \mathcal{E}), \mathcal{D}$). Like that for many Bayesian models, exact inference for our proposed model is difficult due to the intractable partition functions. We therefore exploit the collapsed Gibbs sampling method (a widely used Markov chain Monte Carlo algorithm): first, iteratively sample latent variables (i.e. $\boldsymbol{z}$, $\boldsymbol{c}$, $\boldsymbol{c'}$, $\boldsymbol{s}$, $\boldsymbol{s'}$) such that whose empirical distribution converges to the posterior; then, use those samples to estimate the model parameters of interest. The Gibbs sampling process at each iteration is described as follows.

- Sample community indicators s_e, s'_e for each link $e = (u, v) \in \mathcal{E}$ according to

$$P(s_e = c, s'_e = c' | \mathbf{s_{-e}}, \mathbf{s'_{-e}}, \cdot) \propto \begin{cases} \frac{n_u^{(c)} + \rho}{n_u^{(\cdot)} + C\rho} \cdot \frac{n_v^{(c')} + \rho}{n_v^{(\cdot)} + C\rho} \cdot \frac{n_{c,c'} + \lambda_1}{n_{c,c'} + \lambda_0 + \lambda_1} & l_c \le l_{c'} \\ 0 & l_c > l_{c'} \end{cases} \tag{1}$$

[4] Here, Cat and Mul stand for *Categorical* distribution and *Multinomial* distribution, respectively.

- Sample community indicators c_d, c'_d and the latent topic z_d for each retweet $d = (u, v, \boldsymbol{w}) \in \mathcal{D}$ according to

$$P(c_d = c, c'_d = c' | \mathbf{c_{-d}}, \mathbf{c'_{-d}}, \mathbf{z_{-d}}, z_d = k, \cdot)$$
$$\propto \begin{cases} \dfrac{n_u^{(c)} + \rho}{n_u^{(\cdot)} + C\rho} \cdot \dfrac{n_v^{(c')} + \rho}{n_v^{(\cdot)} + C\rho} \cdot \dfrac{n_c^{(k)} + \alpha}{n_c^{(\cdot)} + K\alpha} & \\ \cdot \dfrac{n_{c'}^{\prime(k)} + \alpha'}{n_{c'}^{\prime(\cdot)} + K\alpha'} \cdot \dfrac{n_{c,c'} + \lambda_1}{n_{c,c'} + \lambda_0 + \lambda_1} & l_c \leq l_{c'} \\ 0 & l_c > l_{c'} \end{cases} \tag{2}$$

$$P(z_d = k | \mathbf{c_{-d}}, \mathbf{c'_{-d}}, \mathbf{z_{-d}}, \mathbf{w}, c_d = c, c'_d = c', \cdot)$$
$$\propto \frac{n_c^{(k)} + \alpha}{n_c^{(\cdot)} + K\alpha} \cdot \frac{n_{c'}^{\prime(k)} + \alpha'}{n_{c'}^{\prime(\cdot)} + K\alpha'} \cdot \frac{n^{(k)} + \beta}{n^{(\cdot)} + K\beta} \cdot \frac{\prod_{v=1}^{V} \prod_{q=0}^{n_d^{(v)}-1} (n_k^{(v)} + q + \gamma)}{\prod_{q=0}^{n_d^{(\cdot)}-1} (n_k^{(\cdot)} + q + V\gamma)} \tag{3}$$

In the equations above, $n_u^{(c)}$ denotes the number of times when user u acts as a member of community c in all links and retweets, $n_c^{(k)}$ denotes the number of posts of topic k retweeted by community c, $n_{c'}^{\prime(k)}$ is the number of posts of topic k retweeted from community c', $n_{c,c'}$ is the number of links from community c to community c' and posts retweeted by community c from community c', $n_k^{(v)}$ represents the number of times word v is assigned to topic k and $n^{(k)}$ represents the number of posts assigned to topic k. $n_d^{(v)}$ is pre-calculated counts denoting the number of times word v occurs in retweet d. We define $\mathbf{z_{-d}}$ as all the topic indicators for posts except z_d; and the same with $\mathbf{s_{-e}}$, $\mathbf{s'_{-e}}$, $\mathbf{c_{-d}}$, and $\mathbf{c'_{-d}}$. All the counts are calculated with link e excluded in Eq. (1) and with post d excluded in Eqs. (2-3).

After an appropriate number of iterations until convergence, the estimates for the latent space parameters can be obtained with the following formulas:

- Community membership π: $\pi_{uc} = \frac{n_u^{(c)}+\rho}{n_u^{(\cdot)}+C\rho}$.
- Topic interests θ: $\theta_{ck} = \frac{n_c^{(k)}+\alpha}{n_c^{(\cdot)}+K\alpha}$.
- Topic popularity θ': $\theta'_{c'k} = \frac{n_{c'}^{\prime(k)}+\alpha'}{n_{c'}^{\prime(\cdot)}+K\alpha'}$.
- Influence strength η: $\eta_{cc'} = \frac{n_{c,c'}+\lambda_1}{n_{c,c'}+\lambda_0+\lambda_1}$.
- Topic distribution $\boldsymbol{\phi}$: $\phi_k = \frac{n^{(k)}+\beta}{n^{(\cdot)}+K\beta}$.
- Word distribution $\boldsymbol{\psi}$: $\psi_{kv} = \frac{n_k^{(v)}+\gamma}{n_k^{(\cdot)}+V\gamma}$.

Note again that, we first assign topic level indicators l_c to each community c according to pre-given ratios for different hierarchical levels, and then implement the hierarchical community discovery in the inference stage under the assumption that no anti-hierarchy links and retweet actions would be observed. In other words, the posterior probability of community indicators c and c' is zero if they violate the hierarchical rules with $l_c > l_{c'}$, as we can see in Eqs. (1-2) so that $\eta_{cc'}$ is minimized. The mathematical derivation and convergence study of our constrained Gibbs sampler are reported in the Appendix.

Model Application. The obtained model parameters can be used in several application problems. Specifically, $\boldsymbol{\pi}$ represents the hierarchical *community structure* discovered by HCID, which can be applied to collaborative recommendation tasks. Also, $\boldsymbol{\phi}$ can be exploited in *topic extraction*, while $\boldsymbol{\eta}$ along with communities' topic interest $\boldsymbol{\theta}$ and topic popularity $\boldsymbol{\theta'}$ provides input parameters for influence propagation models (e.g., Topic-aware Independent Cascade model [3]) to conduct topic-aware *top-K influencer retrieval* at the community level. At last, the major application of our model is *information diffusion prediction*, which will be covered in Section 2.4.

2.4 Prediction Method

2.4.1 Retweet Prediction. Retweet prediction is a common task of diffusion analysis, whose aim is to predict whether a post d with textual content $\mathbf{w_d}$ from user v will be retweeted by another user u in microblog settings. In other words, we are asked to estimate the conditional probability that user u will retweet a post given its content $\mathbf{w_d}$ as well as source user v, namely $P(u|\mathbf{w_d}, v)$[5].

In our approach, we regard this probability as a mixture over the topic-aware user-to-user influence probabilities $P(u|k, v)$ for all the topics $k \in [1, 2, ..., K]$,

$$P(u|\mathbf{w_d}, v) = \sum_k P(k|\mathbf{w_d}, v) P(u|k, v), \tag{4}$$

with mixture weights inferred through post d's content and its source user v's topic interest,

$$P(k|\mathbf{w_d}, v) \propto P(k|v) P(\mathbf{w_d}|k, v) = \prod_l \psi_{k w_{dl}} \cdot \sum_c \pi_{vc} \theta_{ck}. \tag{5}$$

The user-to-user influence can be converted into community-level diffusion probabilities via community memberships,

$$P(u|k, v) = \sum_{c,c'} \pi_{uc} \cdot \pi_{vc'} \cdot P(c|k, c'), \tag{6}$$

while the latter can be decomposed to the product of topic-irrelevant influence strengths $\boldsymbol{\eta}$, target communities' topic interest $\boldsymbol{\theta}$ and source communities' topic popularity $\boldsymbol{\theta'}$ as discussed in Section 2.2

$$P(c|k, c') = \eta_{cc'} \cdot \theta_{ck} \cdot \theta'_{c'k}. \tag{7}$$

As in [19], we also find that top few communities are sufficient to represent users' behavioral preferences. On average, the top three community memberships of a user u add up to 85.2%, and the top ten 95.1%. This enables us to save a lot of computational costs by only considering users' top communities c and c' in Eqs. (5-6), while top communities of each user can be obtained easily in the offline pre-processing stage.

Compared to similar work, our prediction method not only considers users' topic interests, but also takes source users' *topic popularity* into account, which well captures users' topic expertise, topic-aware authoritativeness and other likewise phenomena. Additionally, influence strengths (i.e., η) not only indicates social tie strengths, but also reflect social hierarchy in information diffusion.

[5]Here, we slightly abuse the probability notation.

2.4.2 Cascade Prediction. In the retweet prediction, we focus on direct user-to-user retweet behaviors, which serve as local "hops" in the whole information diffusion process. In other words, the content $\mathbf{w_d}$ of a post retweeted by user u_{target} from user u_{source} may not be originated by u_{source}; it can also be a post retweeted from u_{source} but originated from u_{root}. In the latter case, we say that u_{target}, u_{source} and other users who have retweeted the same content are contaminated users in the information cascade of content $\mathbf{w_d}$ published by u_{root}. As we can see, the information cascading is global information diffusion phenomenon. In this task, our goal is to predict whether a user u will be contaminated in a given information cascade $q = (u_{root}, \mathbf{w_d})$.

We first study the cascading problem at the community level. For each topic k, we construct a graph of communities $G^{(k)} = (C, A^{(k)})$, where A^k is an adjacency matrix with each component $a^{(k)}_{c,c'} = P(c|k, c')$, topic-aware community-level diffusion probability given in Eq. (7). Then, we define the probability that community c will be contaminated by a cascade originated from c' as

$$P_c(c|k, c') = \max_{(c_1=c, c_2, \ldots, c_m=c')\in C^*} \prod_{i\in[1,2,\ldots,m-1]} a^{(k)}_{c_i,c_{i+1}}, \quad (8)$$

that is, the largest influence community c' can have on community c through any possible path. This can be implemented easily with the Floyd-Warshall algorithm [10] by taking negative logarithm of edge weights and finding shortest paths for each pair of communities. Note that, it is both theoretically unreasonable and computationally prohibitive for traditional user-level models to utilize this approach due to the huge differences among users' in-degrees and the large number of users.

At last, we use the same method as in the retweet prediction to obtain the probability $P_c(u|\mathbf{w_d}, v)$ of user u being contaminated by substituting $P(c|k, c')$ in Eq. (6) with $P_c(c|k, c')$.

2.5 Complexity Analysis

Since only positive links and retweet logs are modeled in HCID, the time complexity of the Gibbs sampling inference is linear w.r.t the size of input data and times of iterations T, i.e., $O(T(|\mathcal{E}| + |\mathcal{D}|))$ given parameters C and K fixed. Further, in each iteration, the sampling time for community indicators associated with each link and each retweet is $O(C^2)$, while the sampling for topics is $O(|\boldsymbol{w_d}| \cdot K)$, where $|\mathbf{w_d}|$ is the number of words in retweet d usually with an upper limit of hundreds of words in microblogs.

The complexity of information diffusion prediction consists of two parts. The offline pre-processing stage for retweet prediction is $O(UC)$ for extracting top-k communities for each user, while the cost for cascade prediction is $O(C(U + C^2))$ for an additional run of the Floyd-Warshall algorithm. After that, the online prediction for each retweet action takes only $O(|\boldsymbol{w}| \cdot K)$ in both of the two tasks.

3 EXPERIMENTS

In this section, we evaluate our proposed model using a real Weibo dataset. The experimental results show that our proposed model achieves substantial predictive performance improvement. We also conduct an analysis to uncover why it outperforms other baseline methods.

In the following experiments, we only study two-layered community hierarchy for a simple and clear demonstration, although community hierarchy with multiple levels can be implemented with no extra cost. We use h to denote the ratios of higher level communities. Further empirical study on community hierarchy is reported at the end of this section.

3.1 Experimental Setup

3.1.1 Dataset. **Weibo.** We use a real-world dataset of Sina Weibo from [37], containing both a relationship network $G = (\mathcal{U}, \mathcal{E})$ and retweet logs $\mathcal{D} = \{(u, v, \boldsymbol{w_d})\}$.

In the data preprocessing step, stop words and the words occurred only once in the corpus are excluded; inactive users associated with less than 20 links and retweet logs in total are removed. All the retweet logs were made between June 1, 2012 and August 31, 2012. In the final dataset we obtain, there are about 68K users, 8.1M links and 2.7M retweet logs containing 110M words in total.

The retweet logs $\mathcal{D}$ are split into training $\mathcal{D}_{training}$(80%) and test $\mathcal{D}_{test}$(20%). Although only true positive samples are modeled in our approach, we need negative samples to evaluate predictive performance of our model. We therefore construct a half-sized negative test set $\overline{\mathcal{D}_{test}}$ randomly selected from $\overline{\mathcal{D}} = \{(u_{neg}, v, \mathbf{w_d}) \notin \mathcal{D}\}$, where u_{neg} is a follower of v who never retweeted a post with content $\mathbf{w_d}$ from v.

Moreover, to test model performance in the cascade prediction task, we generate a cascade set $Q_{test} = \{(u_{root}, \boldsymbol{w}, U_c, \overline{U_c})\}$, where U_c is the set of contaminated users who retweeted the original post from u_{root} with content $\boldsymbol{w}$ while $\overline{U_c}$ is the set of randomly sampled uncontaminated users. It contains about 2K original posts' cascading records which never appear in $\mathcal{D}_{training}$, with average 19.6 contaminated users within each cascade.

DBLP. We further validate our model on an academic social network from DBLP[32]. The dataset is generated in a similar manner to the Weibo dataset. Co-authorships and citations are regarded respectively as "friendships" and "retweets", where "tweets" are titles along with abstracts of the cited research papers. In summary, there are about 21K users, 2.3M links, and 5.9M "tweets" containing 412M words in total.

3.1.2 Baselines. We conduct comparison study with the following baseline models. Model (1-2) are two state-of-art information diffusion modeling methods used to demonstrate performance improvement achieved by our approach. Model (3-4) are variants of our proposed model used to illustrate the contribution of two notions proposed in this paper, i.e., *community hierarchy* and *topic popularity*.

(1) **COLD.** Proposed in [19], COLD (COmmunity Level Diffusion) is the state-of-art generative model for community level information diffusion. In COLD, each user is also represented by a community membership distribution, and each community has a topic distribution that generates all the posts published by itself. An interaction network derived from retweeting logs without semantic information is modeled by pairwise Bernoulli distributions.
(2) **CDK.** Proposed in [6], CDK (Content Diffusion Kernel) is a network embedding method to represent users in a latent vector space in such a way that information diffusion can

Table 2: AUC values of retweet prediction on Weibo. C and K are the number of communities and topics, respectively. Results of our proposed model are given for several values of h, the ratio of the higher level communities. Results of the baseline method *No-Popularity* are given for the best choice of h.

Method	K	C=50	C=100	C=150
COLD	40	0.662	0.688	0.679
	60	0.698	0.663	0.669
No-Popularity	40	0.750	0.788	0.754
	60	0.746	0.770	0.776
No-Hierarchy	40	0.812	0.827	0.820
	60	0.800	0.817	0.831
HCID (h = 10%)	40	0.824	0.830	0.821
	60	0.813	0.828	0.821
HCID (h = 25%)	40	**0.833**	**0.842**	**0.846**
	60	0.828	0.834	0.841
HCID (h = 50%)	40	0.829	**0.842**	0.844
	60	0.827	0.840	0.843

be regarded as a heat diffusion process in that space. Information diffusion between two users is thereby predicted by their Euclidean distance.

(3) **No-Popularity.** In this variant, we only use one topic distribution to represent each community's topical preference, thus not capable of differentiating communities' topic popularity from their topic interests.

(4) **No-Hierarchy.** In this variant, community hierarchy is not considered, and therefore links and retweet interactions can be established by any two communities.

3.2 Information Diffusion Prediction

We evaluate HCID's predictive performance of information diffusion in two sub-tasks, the retweet prediction and the cascade prediction. The two tasks respectively represent one "hop" and the whole process of information diffusion.

3.2.1 Retweet Prediction Performance. The problem formulation of retweet prediction is given in Section 2.4. Since no pre-defined threshold for the existence of a given post in both our proposed model and baseline models, we regard the diffusion prediction as an information retrieval problem aiming to rank $\mathcal{D}_{test}$ before $\overline{\mathcal{D}_{test}}$ in terms of the diffusion probability calculated in Eq. (4), and use Area Under the ROC Curve (AUC) and Precision-Recall Curves as metrics. Note that CDK is not applicable to this task so that we only compare the other baseline methods.

Table 2 lists the results of all comparison approaches on the Weibo dataset. We can see that all three baseline methods perform worse than our proposed model HCID. Figure 5 visualizes their performances using Precision-Recall curves. We observe that more topics do not necessarily bring better performances. It is possibly because users usually do not differentiate topics very meticulously when making retweet decisions and the overestimated number of topics may cause overfitting problems.

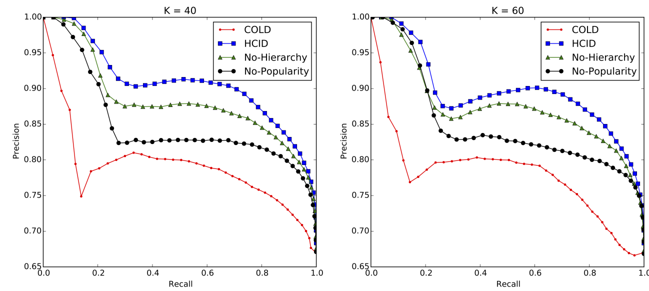

Figure 5: Retweet Prediction Performance on Weibo. h is set to the optimal options for HCID.

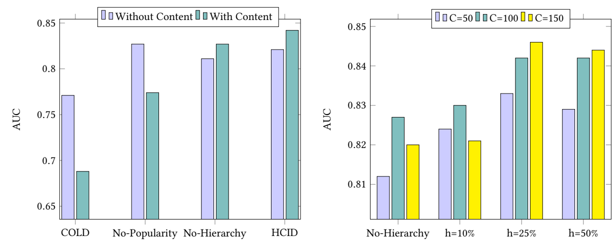

Figure 6: Performance Improvement Analysis on Weibo. If not specified, C, K and h are set to the optimal parameter options, i.e., 100, 40, and 25%, respectively.

We conduct further analysis on the performance improvement by HCID. As discussed in Section 2.2, a retweet decision depends on two parts, the topic-irrelevant influence strength from the source user to the target and the content of the post. Figure 6(left) shows the predictive performance results without and with the latter part, respectively. We can see that the semantic content helps for No-Hierarchy and HCID to obtain better performance, while the opposite for COLD and No-Popularity. This is not surprising for COLD and No-Popularity which use the same distributions to model users' topic preferences both as information spreaders and as information receivers. In fact, there is only an averaged Jaccard similarity of 0.486 between the two sets of top three topics that a community is most interested in and has the most popularity in, respectively. This number is even smaller for lower level communities, which is only 0.342, possibly because lower level communities generally have diverse interests and unpredictable popularity in those topics. Besides, without taking advantage of posts' retweeting source makes COLD exhibit even worse performance. Furthermore, we can find that HCID achieves better performance than No-Hierarchy mainly because the introduction of community hierarchy improves the community structure extraction, hence the estimation of topic-irrelevant influence strengths.

Figure 6(right) shows the impact of different ratios of higher communities h on the information diffusion prediction. For small h (e.g., h = 10%), it is sufficient to use 100 communities to group users who connect to each others closely and share similar behavior preferences. On the other hand, as h increases, users can be efficiently

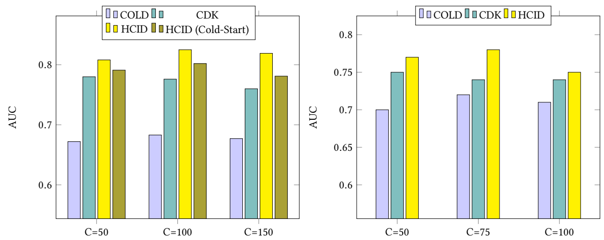

Figure 7: Cascade Prediction Performance on the Weibo(left) and DBLP(right) dataset. C denotes the latent space dimensions for CDK, and h is set to the optimal options for HCID.

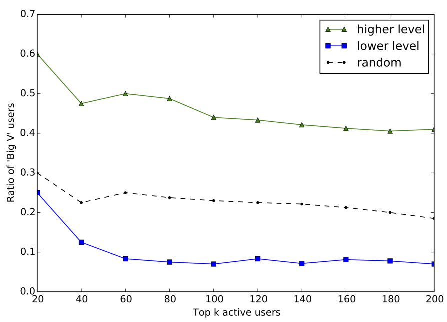

Figure 8: Comparison of the ratios of "Big V"s in active users in different sets of communities in Weibo.

divided into more finer-grained communities with more precise retweet patterns, thus achieving better predictive performance.

Therefore, we can conclude that the substantial performance improvement by our proposed model HCID is achieved by incorporating the notions of *topic popularity* and *community hierarchy*, along with the holistic retweet modeling approach taking posts' retweeting source into account.

3.2.2 Cascade Prediction Performance. In addition to the capability of predicting a single "hop" (i.e., a retweet interaction between two users) in the information diffusion process, our model HCID is able to predict the whole message cascade. In order to evaluate the prediction results, we treat the prediction for each cascade $q \in Q_{test}$ as a retrieval task whose goal is to rank contaminated users U_c before the uncontaminated $\overline{U_c}$ in that cascade. We then average the AUC values for all cascades.

It is worthwhile to mention that there are some users (about 5%) in Weibo test set Q_{test} who never posted any message recorded in the retweet logs $D_{training}$. Neither COLD nor CDK can predict retweet actions regarding those users because no community membership distribution or embedding vector exists for them. Nonetheless, our model HCID can take advantage of users' friendship network, and cold-start users associated with no retweet logs can still be given a high-quality community membership distribution collaboratively. In our experiment, we first remove cold-start users for a fair comparison. After that, we evaluate our model on the complete test. As shown in Figure 7, HCID outperforms other baseline methods consistently on both Weibo and DBLP datasets.

Table 3: Comparison of the statistics of higher and lower level communities with h = 25% and C = 100.

Statistics	Higher level	Lower level
ratio of participation (tol.)	23.72%	76.28%
#user per community (avg.)	6372.4	6980.1
pagerank (avg.)	1.288	0.707
in-degree centrality (avg.)	0.0170	0.0078
betweenness centrality (avg.)	0.00020	0.00009

3.3 Community Hierarchy

Experiments above show that introducing the notion of community hierarchy can bring substantial improvement with almost no additional cost. In this section, we will give an empirical comparison of higher and lower communities for more insights.

Table 3 demonstrates some statistics of higher and lower communities in Weibo. Although social users seem to participate equally in both of the higher and lower level communities, various metrics indicate users who are actively involved in higher level communities have higher centrality than the others. As mentioned in Section 1, those higher level communities are more likely to be venues for influential users who seldom follow "ordinary people" in the lower level and cautiously retweet their posts due to high social status and concerns about public opinions. Figure 8 justifies this interpretation. The ratio of "Big V"s[6] among active users in higher level communities is significantly larger than that in lower level communities.

Similar results can be also obtained in DBLP. It is worth mentioning that higher level community members (h is set to 25%) are up to 7 times more likely to serve as program committee (PC) members of top-tier conferences, which shows that the community hierarchy can reveal seniority and general authoritativeness (compared to topical authoritativeness) of researchers in citation networks. The conference data is obtained from [23], which covers six mainstream research areas in computer science.

4 RELATED WORK

Our work is closely related to information diffusion and probabilistic semantic analysis, while existing work on community detection inspires our idea as well.

Information Diffusion. Information diffusion is a vast research domain, attracting extensive research interests [15], including influence maximization (e.g. [20, 21]) and influential node detection (e.g. [23, 27]) with two types of information propagation models, Independent Cascade Model [11] and Linear Threshold Model [14]. Behind those approaches, there is a fundamental problem to model and estimate social influence between two nodes/users (e.g. [12, 13, 33]), which is also more relevant to our work.

Gomez-Rodriguez et. al. proposes NETINF [12] to infer influence strengths between news media sites and blogs from citation cascades. Tang et. al. [31] leverages a factor graph model to investigate the relationship between conformity and influence with retweet cascades and users' attributes. Topical Affinity Propagation (TAP) [30] models topic-level social influence, but it regards

[6] In Weibo, verified accounts with more than 0.5M followers are called "Big V", a well-recognized concept to represent influential users.

topic extraction only as a pre-processing stage without considering topics' interdependence with network structures. Barbieri et. al. [3] also incorporates topics into diffusion models and proposes estimation methods for topic-aware influence strengths. However, the topics are totally based on network structure without taking advantage of semantic information. Finally, our work is most related to the Community Level Diffusion (COLD) model, which is a novel approach to model information diffusion at the community level proposed by Hu et. al. [19].

Probabilistic Semantic Analysis. Another line of existing research is the probabilistic semantic analysis. Blei et. al. first proposes a generative topic model LDA [5]. After that, Cha et. al. [7] apply topic models to social network analysis. Zhao et. al. [39] propose a content-based user in microblog settings.

Furthermore, various topic models are proposed combining semantic content with linking structures both in document citation networks (e.g. Topic-Link-LDA [22], RTM [8]) and in social networks (e.g. SRTM [16], FLDA [4]). In most of those models, textual contents and links are assumed to be generated by the same topic distribution (in some literature, it is also called mixed-membership of communities). However, this assumption is not well fitted with social networks, especially microblogs, in two major reasons: first, links are not necessarily relevant to users' topic interests and they also establish links due to social relationships or conformity; second, users' strong interests in a certain topic do not always indicate that they have significant popularity in that topic as well and such complexity cannot be captured by a single consistent interest distribution for each user.

Community Detection. Our model captures social relationships and users' preferences at the community level and can be applied to community detection as well. While extensive discriminative approaches are proposed [26], our work is more relevant to probabilistic generative community detection models. Airoldi et. al. proposes the Mixed Membership Stochastic Block (MMSB) model [1] where each user is associated with a membership distribution over communities as HCID. Bonchi et. al. proposes the Cascade-based Community Detection (CCN) model [2], which discovers overlapping communities by leveraging information cascading behaviors. Han et. al. introduces users' social roles into network structure modeling and community detection in the Community Role Model (CRM) [17], which inspires our idea of community hierarchy to some degree.

5 CONCLUSION

In this paper, we propose a Hierarchical Community-level Information Diffusion (HCID) model. This model can not only help us better understand information diffusion patterns by extracting communities which well reflect both the *structural* and the *semantic* dimensions of information diffusion, but also have strong predictive capability. We evaluate our model on two real datasets from, Weibo and DBLP. Results show that our model outperforms other baseline approaches, thanks to two innovative notions *community hierarchy* and *topic popularity* that we propose in this paper.

There are several ways to extend our work in the future. One of them is deeper investigation on more complex interaction mechanisms across multiple hierarchical community levels. We also plan to incorporate more types of social contagions, such as clicking "like", into our generative framework to better understand information diffusion processes. Lastly, it might be an interesting and challenging problem to study information diffusion across dynamic communities with changing hierarchical structures.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their valuable comments and helpful suggestions. This work is supported by 973 Program under Grant No. 2014CB340405, NSFC under Grant No. 61532001 and No.61370054, and MOE-RCOE under Grant No. 2016ZD201.

REFERENCES

[1] Edoardo M Airoldi, David M Blei, Stephen E Fienberg, and Eric P Xing. 2008. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research* 9, Sep (2008), 1981–2014.

[2] Nicola Barbieri, Francesco Bonchi, and Giuseppe Manco. 2013. Cascade-based community detection. In *Proceedings of the sixth ACM international conference on Web search and data mining*. ACM, 33–42.

[3] Nicola Barbieri, Francesco Bonchi, and Giuseppe Manco. 2013. Topic-aware social influence propagation models. *Knowledge and Information System* 37, 3 (2013), 555–584.

[4] Bin Bi, Yuanyuan Tian, Yannis Sismanis, Andrey Balmin, and Junghoo Cho. 2014. Scalable Topic-specific Influence Analysis on Microblogs. In *Proceedings of the 7th ACM international conference on Web search and data mining*. ACM, 513–522.

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

[6] Simon Bourigault, Cedric Lagnier, Sylvain Lamprier, Ludovic Denoyer, and Patrick Gallinari. 2014. Learning social network embeddings for predicting information diffusion. In *Proceedings of the 7th ACM international conference on Web search and data mining*. ACM, 393–402.

[7] Youngchul Cha and Junghoo Cho. 2012. Social-network analysis using topic models. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 565–574.

[8] Jonathan Chang and David M Blei. 2009. Relational topic models for document networks. In *AISTATS*. 81–88.

[9] Elena Erosheva, Stephen Fienberg, and John Lafferty. 2004. Mixed-membership models of scientific publications. *Proceedings of the National Academy of Sciences* 101 (2004), 5220–5227.

[10] Robert W. Floyd. 1962. Algorithm 97: Shortest Path. *Commun. ACM* 5, 6 (June 1962), 345–. DOI: http://dx.doi.org/10.1145/367766.368168

[11] Jacob Goldenberg, Barak Libai, and Eitan Muller. 2001. Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth. *Marketing Letters* 12, 3 (2001), 211–223.

[12] Manuel Gomez Rodriguez, Jure Leskovec, and Andreas Krause. 2010. Inferring networks of diffusion and influence. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1019–1028.

[13] Amit Goyal, Francesco Bonchi, and Laks VS Lakshmanan. 2010. Learning influence probabilities in social networks. In *Proceedings of the third ACM international conference on Web search and data mining*. ACM, 241–250.

[14] Mark Granovetter. 1978. Threshold Models of Collective Behavior. *Amer. J. Sociology* 83, 6 (1978), 1420–1443.

[15] Adrien Guille, Hakim Hacid, Cécile Favre, and Djamel A Zighed. 2013. Information diffusion in online social networks: A survey. *SIGMOD Record* 42, 2 (2013), 17–28.

[16] Weiyu Guo, Shu Wu, Liang Wang, and Tieniu Tan. 2015. Social-Relational Topic Model for Social Networks. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 1731–1734.

[17] Yu Han and Jie Tang. 2015. Probabilistic community and role model for social networks. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 407–416.

[18] Qirong Ho, Rong Yan, Rajat Raina, and Eric P. Xing. 2012. Understanding the Interaction between Interests, Conversations and Friendships in Facebook. *CoRR* abs/1211.0028 (2012).

[19] Zhiting Hu, Junjie Yao, Bin Cui, and Eric Xing. 2015. Community level diffusion extraction. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*. ACM, 1555–1569.

[20] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 137–146.

[21] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. 2007. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 420–429.

[22] Yan Liu, Alexandru Niculescu-Mizil, and Wojciech Gryc. 2009. Topic-link LDA: joint models of topic and author community. In *proceedings of the 26th annual international conference on machine learning*. ACM, 665–672.

[23] Tiancheng Lou and Jie Tang. 2013. Mining structural hole spanners through information diffusion in social networks. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, 825–836.

[24] Ramesh Nallapati and William W Cohen. 2008. Link-PLSA-LDA: A New Unsupervised Model for Topics and Influence of Blogs. In *Proceedings of the International Conference on Web and Social Media*.

[25] Mark EJ Newman. 2002. Assortative mixing in networks. *Physical review letters* 89, 20 (2002), 208701.

[26] Symeon Papadopoulos, Yiannis Kompatsiaris, Athena Vakali, and Ploutarchos Spyridonos. 2012. Community detection in social media. *Data Mining and Knowledge Discovery* 24, 3 (2012), 515–554.

[27] Mojtaba Rezvani, Weifa Liang, Wenzheng Xu, and Chengfei Liu. 2015. Identifying top-k structural hole spanners in large-scale social networks. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 263–272.

[28] Wilbur Schramm and Porter William. 1982. *Men, women, messages, and media: Understanding human communication*. New York: Harper & Row.

[29] Mohsen Shahriari, Sebastian Krott, and Ralf Klamma. 2015. Disassortative Degree Mixing and Information Diffusion for Overlapping Community Detection in Social Networks (DMID). In *Proceedings of the 24th International Conference on World Wide Web Companion*. 1369–1374.

[30] Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. 2009. Social influence analysis in large-scale networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 807–816.

[31] Jie Tang, Sen Wu, and Jimeng Sun. 2013. Confluence: Conformity influence in large social networks. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 347–355.

[32] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. 2008. Arnetminer: extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 990–998.

[33] Rongjing Xiang, Jennifer Neville, and Monica Rogati. 2010. Modeling relationship strength in online social networks. In *Proceedings of the 19th international conference on World wide web*. ACM, 981–990.

[34] Yang Yang, Jie Tang, Cane Wing-ki Leung, Yizhou Sun, Qicong Chen, Juanzi Lit, and Qiang Yang. 2015. RAIN: Social Role-aware Information Diffusion. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. AAAI, 367–373.

[35] Zi Yang, Jingyi Guo, Keke Cai, Jie Tang, Juanzi Li, Li Zhang, and Zhong Su. 2010. Understanding retweeting behaviors in social networks. In *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, 1633–1636.

[36] Nicholas Jing Yuan, Yuan Zhong, Fuzheng Zhang, Xing Xie, Chin Yew Lin, and Yong Rui. 2016. Who Will Reply to/Retweet This Tweet?: The Dynamics of Intimacy from Online Social Interactions. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*. ACM, 3–12.

[37] Jing Zhang, Jie Tang, Juanzi Li, Yang Liu, and Chunxiao Xing. 2015. Who influenced you? predicting retweet via social influence locality. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 9, 3 (2015), 25.

[38] Wayne Xin Zhao, Jing Jiang, Jianshu Weng, Jing He, Ee-Peng Lim, Hongfei Yan, and Xiaoming Li. 2011. Comparing twitter and traditional media using topic models. In *European Conference on Information Retrieval*. Springer, 338–349.

[39] Yukun Zhao, Shangsong Liang, Zhaochun Ren, Jun Ma, Emine Yilmaz, and Maarten de Rijke. 2016. Explainable user clustering in short text streams. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 155–164.

A SAMPLING HIERARCHICAL COMMUNITY INDICATORS

In this paper, we design a constrained Gibbs sampling algorithm under the assumption that no anti-hierarchy links and retweets would be observed. We denote this assumption as $\mathcal{A} = \bigcup_{e \in \mathcal{E}} \mathcal{A}_e \cup \bigcup_{d \in \mathcal{D}} \mathcal{A}_d$, where $\mathcal{A}_e$ and $\mathcal{A}_d$ represent the event $l_{s_e} \leq l_{s'_e}$ and $l_{c_d} \leq l_{c'_d}$ respectively. Now we show the derivation of Eqs. (1-2).

First, using Bayes' rule, we can get the joint distribution of all latent variables under the assumption $\mathcal{A}$

$$\begin{aligned} &P(\boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s}, \boldsymbol{s'}, z|\mathcal{A}, \cdot) \\ &\quad \propto P(\mathcal{A}|\boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s}, \boldsymbol{s'}, z, \cdot) \cdot P(\boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s}, \boldsymbol{s'}, z|, \cdot). \end{aligned} \tag{9}$$

From the graphical model representation in Figure 4, we then obtain

$$\begin{aligned} &P(\boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s}, \boldsymbol{s'}, z|\cdot) \\ &\propto P(\boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s}, \boldsymbol{s'}|\rho) \cdot P(z|\beta) \cdot P(\boldsymbol{w}|z, \gamma) \cdot P(\boldsymbol{I}, \boldsymbol{I'}|\boldsymbol{\lambda}, \boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s}, \boldsymbol{s'}). \end{aligned} \tag{10}$$

The conditional of s_e and s'_e can be calculated by dividing the joint distribution of all latent variables by the joint of all variables except s_e and s'_e,

$$\begin{aligned} &P(s_e = c, s'_e = c'|\boldsymbol{s_{-e}}, \boldsymbol{s'_{-e}}, \boldsymbol{c}, \boldsymbol{c'}, z, \mathcal{A}, \cdot) \\ &= \frac{P(\boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s}, \boldsymbol{s'}, z|\mathcal{A}, \cdot)}{P(\boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s_{-e}}, \boldsymbol{s'_{-e}}, z|\mathcal{A}, \cdot)} \\ &\propto \frac{P(\mathcal{A}|\boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s}, \boldsymbol{s'}, z, \cdot)}{P(\mathcal{A}_{-e}|\boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s_{-e}}, \boldsymbol{s'_{-e}}, z, \cdot)} \cdot \frac{P(\boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s}, \boldsymbol{s'}, z|\cdot)}{P(\boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s_{-e}}, \boldsymbol{s'_{-e}}, z|\cdot)} \\ &\propto P(\mathcal{A}_e|s_e = c, s'_e = c') \cdot \frac{P(\boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s}, \boldsymbol{s'}|\rho)}{P(\boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s_{-e}}, \boldsymbol{s'_{-e}}|\rho)} \\ &\cdot \frac{P(\boldsymbol{I}, \boldsymbol{I'}|\boldsymbol{\lambda}, \boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s}, \boldsymbol{s'})}{P(\boldsymbol{I}, \boldsymbol{I'_{-e}}|\boldsymbol{\lambda}, \boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s_{-e}}, \boldsymbol{s'_{-e}})}. \end{aligned} \tag{11}$$

The first fraction of Eq. (11) is easy to obtain by the definition of $\mathcal{A}_e$,

$$P(\mathcal{A}_e|s_e, s'_e) = \begin{cases} 1 & l_{s_e} \leq l_{s'_e} \\ 0 & l_{s_e} > l_{s'_e}. \end{cases} \tag{12}$$

Next, we show how to derive the second fraction

$$\frac{P(\boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s}, \boldsymbol{s'}|\rho)}{P(\boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s_{-e}}, \boldsymbol{s'_{-e}}|\rho)} = \frac{\int P(\boldsymbol{\pi}|\rho)P(\boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s}, \boldsymbol{s'}|\boldsymbol{\pi})d\boldsymbol{\pi}}{\int P(\boldsymbol{\pi}|\rho)P(\boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s_{-e}}, \boldsymbol{s'_{-e}}|\boldsymbol{\pi})d\boldsymbol{\pi}}, \tag{13}$$

where the integrals can be calculated in the following way,

$$\begin{aligned} &\int P(\boldsymbol{\pi}|\rho)P(\boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s}, \boldsymbol{s'}|\boldsymbol{\pi})d\boldsymbol{\pi} \\ &= \int \prod_{u^*} \frac{\Gamma(C\rho)}{\prod_c \Gamma(\rho)} \prod_c \pi_{u^*c}^{\rho-1} \cdot \prod_{u^*} \prod_c \pi_{u^*c}^{n_{u^*}^{(c)}} d\boldsymbol{\pi} \\ &= \prod_{u^*} \frac{\Gamma(C\rho)}{\prod_c \Gamma(\rho)} \cdot \frac{\prod_c \Gamma(n_{u^*}^{(c)} + \rho)}{\Gamma(n_{u^*}^{(\cdot)} + C\rho)}. \end{aligned} \tag{14}$$

By plugging Eq. (14) into Eq. (13) and canceling out the same terms with the trick $\Gamma(x+1) = x\Gamma(x)$, we can obtain

$$\frac{P(\boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s}, \boldsymbol{s'}|\rho)}{P(\boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s_{-e}}, \boldsymbol{s'_{-e}}|\rho)} = \frac{n_{u,-e}^{c_e} + \rho}{n_{u,-e}^{(\cdot)} + \rho} \cdot \frac{n_{v,-e}^{c'_e} + \rho}{n_{v,-e}^{(\cdot)} + \rho}, \tag{15}$$

while the third fraction in Eq. (11) can be obtained in a similar fashion,

$$P(\boldsymbol{I}, \boldsymbol{I'_{-e}}|\boldsymbol{\lambda}, \boldsymbol{c}, \boldsymbol{c'}, \boldsymbol{s_{-e}}, \boldsymbol{s'_{-e}}) = \frac{n_{c,c',-e} + \lambda_1}{n_{c,c',-e} + \lambda_0 + \lambda_1}. \tag{16}$$

Plugging Eq. (12) and Eqs. (15-16) into Eq. (11), we can finally derive the sampling formula in Eq. (1), where the assumption $\mathcal{A}$ and subscript $-e$ are omitted for a clear demonstration. Similarly, Eq. (2) can be derived.