[30] M. Steyvers, P. Smyth, and T. Griffiths. Probabilistic author-topic models for information discovery. In *KDD'04*, pages 306–315, 2004.

[31] Y.-C. Tam and T. Schultz. Correlated latent semantic model for unsupervised lm adaptation. In *ICASSP?07*, volume 4, pages IV–41–IV–44, 2007.

[32] J. Tang, A. Fong, B. Wang, and J. Zhang. A unified probabilistic framework for name disambiguation in digital library. *IEEE TKDE*, 24(6):975–987, 2012.

[33] J. Tang, J. Sun, C. Wang, and Z. Yang. Social influence analysis in large-scale networks. In *KDD'09*, pages 807–816, 2009.

[34] J. Tang, S. Wu, J. Sun, and H. Su. Cross-domain collaboration recommendation. In *KDD'12*, pages 1285–1294, 2012.

[35] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su. Arnetminer: Extraction and mining of academic social networks. In *KDD'08*, pages 990–998, 2008.

[36] K. M. Ting and I. H. Witten. Issues in stacked generalization. *Journal of Artificial Intelligence Research*, 10:271–289, 1999.

[37] O. Uzuner, Y. Juo, and P. Szolovits. Evaluating the state-of-the-art in automatic de-identification. *J Am Med Inform Assoc*, 14(5):550–563, 2007.

[38] S. Wu, Z. Fang, and J. Tang. Accurate product name recognition from user generated content. In *ICDM 2012 Contest*, pages 874–877, 2012.

9. APPENDIX

According to the generative process, we could integrate out the multinomial distributions θ, ϑ, ϕ, because the model only uses conjugate priors [12]. We use Eq. 6 as the example to explain its derivation, as it contains both social context and domain knowledge. First we write the joint probability:

$$P(\mathbf{w}, \mathbf{z}, \mathbf{v}|\alpha, \beta, \eta, \gamma) \propto \int P(\mathbf{z}|(\mathbf{v}, \mathbf{v}'), \vartheta) P(\vartheta|\alpha) d\vartheta \int P(\mathbf{w}|\mathbf{z}, \phi) P(\phi|\pi_T) \prod_{k=1}^{T} P(\pi_{k+1}|\pi_k, \eta_c) P(\pi_1|\beta, \eta) d\phi d\pi \tag{7}$$

The conditional of s_i is obtained by dividing the joint distribution of all variables by the joint with all variables but s_i (denoted by $\mathbf{s}_{-i}$) and canceling factors that do not depend on $\mathbf{s}_{-i}$.

$$P(z_{di}|\mathbf{z}_{-di}, \mathbf{w}, .) = \frac{P(\mathbf{w}, \mathbf{z}, \mathbf{v}|\alpha, \beta, \eta, \gamma)}{P(\mathbf{w}, \mathbf{z}_{di}, \mathbf{v}|\alpha, \beta, \eta, \gamma)} = \frac{\int P(\mathbf{w}|\mathbf{z}, \phi) P(\phi|\pi_T) \prod_{k=1}^{T} P(\pi_{k+1}|\pi_k, \eta_c) P(\pi_1|\beta, \eta) d\phi d\pi}{\int P(\mathbf{w}|\mathbf{z}, \phi) P(\phi|\pi_T) \prod_{k=1}^{T} P(\pi_{k+1}|\pi_k, \eta_c) P(\pi_1|\beta, \eta) d\phi d\pi} \cdot \frac{\int P(\mathbf{z}|(\mathbf{v}, \mathbf{v}'), \vartheta) P(\vartheta|\alpha) d\vartheta}{\int P(\mathbf{z}|(\mathbf{v}, \mathbf{v}'), \vartheta) P(\vartheta|\alpha) d\vartheta} \tag{8}$$

The first fraction of Eq. 8 is responsible for sampling word from topic and the second term is responsible for sampling topic from user (with social relationships). We start with the derivation of the second fraction. Specifically, as $P(\mathbf{z}|(\mathbf{v}, \mathbf{v}', \vartheta)$ and $P(\vartheta|\alpha)$ are a conjugate pair of Multinomial-Dirichlet, we could solve the Multinomial-Dirichlet integral using Gibbs sampling [14]:

$$\int P(\mathbf{z}|(\mathbf{v}, \mathbf{v}'), \vartheta) P(\vartheta|\alpha) d\vartheta = \prod_d \frac{1}{\Delta(\alpha)} \prod_z \vartheta_{vv'z}^{n_{vz}+n_{v'z}+\alpha-1} d\vartheta_{vv'} = \prod_d \frac{\Delta(\vec{n_d}+\alpha)}{\Delta(\alpha)},$$

$$\text{with } \Delta(\alpha) = \frac{\Gamma(\alpha)^T}{\Gamma(T\alpha)} \text{ and } \vec{n_d} = \{n_{vz} + n_{v'z}\}_{z=1}^T$$

where (v, v') denotes a social relationship between user v and v'; n_{vz} and $n_{v'z}$ are two numbers obtained when combining the two distributions θ_v and $\theta_{v'}$. Essentially, in this sampling, we smooth the sampled topic from a user v-specific topic distribution θ_v by the mixture $\vartheta_{vv'}$ of θ_v and $\theta_{v'}$. Accordingly, the second fraction of Eq. 8 can be written as:

$$\frac{\int P(\mathbf{z}|(\mathbf{v}, \mathbf{v}'), \vartheta) P(\vartheta|\alpha) d\vartheta}{\int P(\mathbf{z}_{-di}|(\mathbf{v}, \mathbf{v}'), \vartheta) P(\vartheta|\alpha) d\vartheta} = \frac{\prod_d \frac{\Delta(\vec{n_d}+\alpha)}{\Delta(\alpha)}}{\prod_d \frac{\Delta(\vec{n}_{d,\neg i}+\alpha)}{\Delta(\alpha)}} = \frac{\frac{\Gamma(n_{vz}^{-di}+n_{v'z}+\alpha)}{\Gamma(\sum_{z'}(n_{vz'}^{-di}+n_{v'z'}+\alpha))}}{\frac{\Gamma(n_{vz}^{-di}+n_{v'z}+\alpha-1)}{\Gamma([\sum_{z'}(n_{vz'}^{-di}+n_{v'z'}+\alpha)]-1)}} = \frac{n_{vz_{di}}^{-di} + n_{v'z_{di}} + \alpha}{\sum_z (n_{vz_{di}}^{-di} + n_{v'z_{di}} + \alpha)} \tag{9}$$

Here, we use the identity $\Gamma(x+1) = x\Gamma(x)$; the superscript $^{-di}$ denotes a quantity, excluding the current instance. By further considering a tunable parameter γ to balance the importance between n_{vz} and $n_{v'z}$, we can obtain the first term in Eq. 6. Analogously, we can derive the first fraction of Eq. 8. The difference is that ϕ is sampled from a Dirichlet tree distribution instead of a Dirichlet distribution as that used for sampling topic z. To make it more clear, let us assume that there is a concept path $\{c_{di}^1, \cdots, c_{di}^T\}$ from the root node to the leaf word node (excluding the leaf node). To sample the first level concept c^1, we have

$$P(c_{di}^1|\pi, \beta, \eta) = \frac{\prod_i \frac{\Delta(\vec{m_z}+\alpha)}{\Delta(\beta)}}{\prod_i \frac{\Delta(\vec{m}_{z,\neg di}+\alpha)}{\Delta(\beta)}} = \frac{\frac{\Gamma(m_{zc_{di}^1}^{-di}+W_{c_{di}^1}\beta)}{\Gamma(\sum_{c_s}(m_{zc_s}^{-di}+W_{c_s}\beta))}}{\frac{\Gamma(m_{zc_{di}^1}^{-di}+\beta-1)}{\Gamma([\sum_{c_s}(m_{zc_s}^{-di}+W_{c_s}\beta)]-1)}} = \frac{m_{zc_{di}^1}^{-di} + W_{c_{di}^1}\beta}{\sum_{c_s}(m_{zc_s} + W_{c_s}\beta)} \tag{10}$$

We continue to sample the child node in the concept path, until we get the last internal node in the Dirichlet tree. In this way, we could obtain

$$\prod_{k=1}^{T} \frac{m_{z_{di}c_{di}^k}^{-di} + W_{c_{di}^k}\beta}{\sum_{c_s}(m_{z_{di}c_s}^{-di} + W_{c_s^k}\beta)} \tag{11}$$

Then applying a similar sampling process for word from a topic-specific distribution to that in the standard LDA model, we can obtain the first term in Eq. 6.

$$\frac{m_{c_{di}w_{di}}^{-di} + \eta}{\sum_w m_{c_{di}w}^{-di} + W_c\eta} \tag{12}$$

Finally, by combining Eqs. 9, 11, and 12, we obtain Eq. 6.