Online Supplement

This is the supplementary material to the paper "Integrating Users' Contextual Engagements with Their General Preferences: An Interpretable Followee Recommendation Method".

Appendix A: The Preliminary Study in the Theoretical Foundation Section

Purpose. This preliminary study used the behavioral experimental method to examine the causal relationship between contextual engagement and following behavior.

Participants and Design. This experiment is a 2 (contextual engagement: engaged vs. disengaged) × 12 (replicates of accounts) mixed design with contextual engagement as between-subjects factor and replicates of accounts as within-subjects factor. Three hundred and nineteen users (48% females, Mean age = 24.28, SD = 26.45) who are familiar with OSN platforms (i.e., have used OSN platforms for once in the last week) are recruited with exchange to monetary payment. Thirty-two participants who fail the attention check question are removed, yielding a final sample of 287 participants in the subsequent analyses.

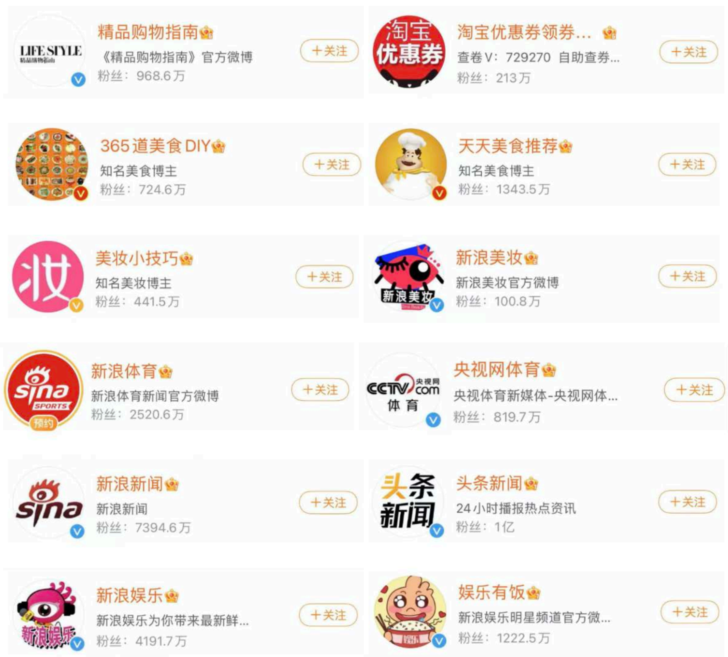


Figure S.1 **Experimental Stimuli of Accounts on "Who-to-Follow"**

Note. The translations of these accounts are: Boutique shopping guide, Taobao coupon, 365 DIY dishes, Daily food recommendation, Beauty tips, Sina beauty makeup, Sina sports, CCTV sports, Sina news, Headline news, Sina entertainment, Active entertainment.

Procedure. After participants admit to participate in this experiment, they are first asked to indicate one of their interests on OSN platforms from six fields, including news, beauty, sports, cooking, shopping, and entertainment. Afterwards, they are randomly and evenly assigned to the engagement and disengagement conditions. Per existing research on consumer behavior (e.g., Donkers et al. 2020), we use the priming task to manipulate participants' contextual engagement. In the engagement condition, participants are asked to write down three reasons why they are interested in this topic. In the disengagement condition, participants just wait approximately the same duration (10 seconds). Then all participants are proceeded to the next task, in which we measure participants' following behaviors. Specifically, participants are asked to imagine that they are selecting 2–3 followees from "who to follow" on Weibo (the most famous OSN platform in China). They are presented with 12 accounts (2 accounts per field; Figure S.1); each account is accompanied with a short description so that participants could easily recognize the topic of these accounts. The order of 12 accounts is randomly presented. Finally, they report their demographics and are thanked. No one suspects the purpose of this experiment.

Result. First, we code participants' choice of accounts that are relevant to their current interest as "1" and choice of accounts that are not relevant to their interest as "−1," and then summarize all their three selected followees as the dependent variable, ranging from −3 to 2. An ANOVA on participants' choice shows that participants in the engagement condition are more likely to choose accounts relevant to their current interest than those in the disengagement condition ($F(1,285) = 3.90$, $p = 0.049$, $\eta_p^2 = .013$). This result is consistent with our rationale that contextual engagement can causally affect users' following behaviors.

Appendix B: Literature Summary on LDA-based Followee Recommendation

Table S.1 Literature on the representative LDA-based Followee Recommendation Methods

Study	Task	Latent Mechanism	Dynamic User Representation	Dataset	
				Type	Details
Pennacchiotti and Gurumurthy (2011)	Followee recommendation	Preference	No	U	Online users' streams
Zhang et al. (2007)	Academic community discovery	Preference	No	S	Academic co-authorship
Henderson and Eliassi-Rad (2009)	Academic community discovery	Preference	No	S	Academic co-authorship
Cha and Cho (2012)	Online community discovery	Preference	No	S	Follower and following lists
Zhao et al. (2013)	Followee recommendation	Preference	No	S	Follower and following lists
Xu et al. (2018)	Followee recommendation	Preference	No	S	Follower and following lists
This study	Followee recommendation	Preference + Engagements	Yes	S	Following list

Note. "U": Unstructured dataset, "S": Structured dataset.

Appendix C: Derivations of Equations (2)–(3).

To obtain Equations (2)–(3), we first have

$$p(z_{u,n},\xi_{u,n}|\mathbf{z}_{\neg(u,n)},\boldsymbol{\xi}_{\neg(u,n)},\mathbf{W})=\frac{p(\mathbf{w}_u,\mathbf{z}_u,\boldsymbol{\xi}_u|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})}{p(\mathbf{w}_u,\mathbf{z}_{\neg(u,n)},\boldsymbol{\xi}_{\neg(u,n)}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})}$$
$$\propto\frac{p(\mathbf{w}_u,\mathbf{z}_u,\boldsymbol{\xi}_u|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})}{p(\mathbf{w}_{\neg(u,n)},\mathbf{z}_{\neg(u,n)},\boldsymbol{\xi}_{\neg(u,n)}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})} \tag{S.1}$$

with the fact that $\mathbf{w}_u$ is observed and thus $p(\mathbf{w}_u)$ is constant.

We repeat Equation (1) (where the hyperparameters are omitted for clarity)

$$p(\mathbf{w}_u,\mathbf{z}_u,\boldsymbol{\xi}_u)=p(w_{u,1},\dots,w_{u,N_u},z_{u,1},\dots,z_{u,N_u},\xi_{u,1},\dots,\xi_{u,N_u})$$
$$=p(w_{u,1}|z_{u,1})\times p(z_{u,1})\times\underbrace{\prod_{n=2}^{N_u}p(w_{u,n}|z_{u,n},\xi_{u,n})\times p(z_{u,n}|z_{u,n-1},\xi_{u,n})\times p(\xi_{u,n})}_{\mathrm{I}}, \tag{S.2}$$

where

$$\underbrace{\prod_{n=2}^{N_u}p(w_{u,n}|z_{u,n},\xi_{u,n})\times p(z_{u,n}|z_{u,n-1},\xi_{u,n})\times p(\xi_{u,n})}_{\mathrm{I}} \tag{S.3}$$
$$=\prod_{\{n:\xi_{u,n}=0\}}p(w_{u,n}|z_{u,n})\times p(z_{u,n}|z_{u,n-1},\xi_{u,n}=0)\times p(\xi_{u,n}=0) \tag{S.4}$$
$$\times\prod_{\{n:\xi_{u,n}=1\}}p(w_{u,n}|z_{u,n})\times p(z_{u,n}|z_{u,n-1},\xi_{u,n}=1)\times p(\xi_{u,n}=1) \tag{S.5}$$

Recall that, as formalized in Algorithm 1, if $\xi_{u,n}=1$, the current interest stays consistent with the previous one and we have $z_{u,n}=z_{u,n-1}$; if $\xi_{u,n}=0$, the current interest is drawn from the general preference and we have $z_{u,n}\sim\text{Multi}(\boldsymbol{\theta}_u)$. By introducing the Multinomial-Dirichlet and Binomial-Beta conjugate, we now derive the factors in Equation (S.5). First, we obtain the word distribution

$$\prod_{\{n:\xi_{u,n}=0\}}\prod_{\{n:\xi_{u,n}=1\}}p(w_{u,n}|z_{u,n})=\int_{\varphi_{z_{u,n}}}\prod_{\{n\}}p(w_{u,n}|\varphi_{z_{u,n}})\times p(\varphi_{z_{u,n}}|\boldsymbol{\beta})d\varphi_{z_{u,n}}$$
$$=\int_{\varphi_{z_{u,n}}}\prod_{t=1}^{V}\varphi_{z_{u,n},t}^{n_t^{(z_{u,n})}}\times\frac{\Gamma(\sum_{t=1}^{V}\beta_t)}{\prod_{t=1}^{V}\Gamma(\beta_t)}\times\prod_{t=1}^{V}\varphi_{z_{u,n},t}^{\beta_t-1}d\varphi_{z_{u,n}}$$
$$=\frac{\Gamma(\sum_{t=1}^{V}\beta_t)}{\prod_{t=1}^{V}\Gamma(\beta_t)}\times\int_{\varphi_{z_{u,n}}}\prod_{t=1}^{V}\varphi_{z_{u,n},t}^{n_t^{(z_{u,n})}+\beta_t-1}d\varphi_{z_{u,n}}$$
$$=\frac{\Gamma(\sum_{t=1}^{V}\beta_t)}{\prod_{t=1}^{V}\Gamma(\beta_t)}\times\frac{\prod_{t=1}^{V}\beta_t}{\prod_{t=1}^{V}\Gamma(\beta_t)}\times\frac{\prod_{t=1}^{V}\Gamma(\beta_t+n_t^{(z_{u,n})})}{\Gamma(\sum_{t=1}^{V}(\beta_t+n_t^{(z_{u,n})}))} \tag{S.6}$$

and the topic distributions given $\xi_{u,n}=0$ and $\xi_{u,n}=1$,

$$\prod_{\{n:\xi_{u,n}=0\}}p(z_{u,n}|z_{u,n-1},\xi_{u,n}=0)=\int_{\boldsymbol{\theta}_u}\prod_{\{n:\xi_{u,n}=0\}}p(z_{u,n}|\boldsymbol{\theta}_u)\times p(\boldsymbol{\theta}_u|\boldsymbol{\alpha})$$
$$=\int_{\boldsymbol{\theta}_u}\prod_{k=1}^{K}\theta_{u,k}^{n_u^{(k)}}\times\frac{\Gamma(\sum_{k=1}^{K}\alpha_k)}{\prod_{k=1}^{K}\Gamma(\alpha_k)}\times\prod_{k=1}^{K}\theta_{u,k}^{\alpha_k-1}d\boldsymbol{\theta}_u$$
$$=\frac{\Gamma(\sum_{k=1}^{K}\alpha_k)}{\prod_{k=1}^{K}\Gamma(\alpha_k)}\times\int_{\boldsymbol{\theta}_u}\prod_{k=1}^{K}\theta_{u,k}^{n_u^{(k)}+\alpha_k-1}d\boldsymbol{\theta}_u$$
$$=\frac{\Gamma(\sum_{k=1}^{K}\alpha_k)}{\prod_{k=1}^{K}\Gamma(\alpha_k)}\times\frac{\prod_{k=1}^{K}\Gamma(\alpha_k+n_u^{(k)})}{\Gamma[\sum_{k=1}^{K}(\alpha_k+n_u^{(k)})]} \tag{S.7}$$

and

$$\prod_{\{n:\xi_{u,n}=1\}} p(z_{u,n}|z_{u,n-1},\xi_{u,n}=1)=1 \tag{S.8}$$

Next, we have

$$\begin{aligned}\prod_{\{n:\xi_{u,n}=0\}} p(\xi_{u,n}=0) &= \int_0^1 \prod_{\{n:\xi_{u,n}=0\}} p(\xi_{u,n}|\psi_u)\times p(\psi_u|\boldsymbol{\gamma})d\psi_u \\ &= \int_0^1 (1-\psi_u)^{n_u^{(0)}} \times \frac{1}{\mathrm{Beta}(\gamma_0,\gamma_1)} \times \psi_u^{\gamma_0-1}\times(1-\psi_u)^{\gamma_1-1}d\psi_u \\ &= \frac{\mathrm{Beta}(\gamma_0,\gamma_1+n_u^{(0)})}{\mathrm{Beta}(\gamma_0,\gamma_1)}\end{aligned} \tag{S.9}$$

and

$$\begin{aligned}\prod_{\{n:\xi_{u,n}=1\}} p(\xi_{u,n}=1) &= \int_0^1 \prod_{\{n:\xi_{u,n}=1\}} p(\xi_{u,n}|\psi_u)\times p(\psi_u|\boldsymbol{\gamma})d\psi_u \\ &= \int_0^1 (\psi_u)^{n_u^{(1)}} \times \frac{1}{\mathrm{Beta}(\gamma_0,\gamma_1)} \times \psi_u^{\gamma_0-1}\times(1-\psi_u)^{\gamma_1-1}d\psi_u \\ &= \frac{\mathrm{Beta}(\gamma_0+n_u^{(1)},\gamma_1)}{\mathrm{Beta}(\gamma_0,\gamma_1)}\end{aligned} \tag{S.10}$$

By integrating Equation (S.1) to Equation (S.10) and using $\Gamma(x+1)=x\Gamma(x)$ and $\mathrm{Beta}(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$, we obtain Equations (2) and (3),

$$\begin{aligned}&p(z_{u,n}=k,\xi_{u,n}=1|\mathbf{z}_{\neg(u,n)},\boldsymbol{\xi}_{\neg(u,n)},\mathbf{W}) \\ &\quad= \frac{\prod_{\{n:\xi_{u,n}=1\}} p(w_{u,n}|z_{u,n})\times p(z_{u,n}|z_{u,n-1},\xi_{u,n}=1)\times p(\xi_{u,n}=1)}{\prod_{\{n:\xi_{u,n}=1\}\neg(u,n)} p(w_{u,n}|z_{u,n})\times p(z_{u,n}|z_{u,n-1},\xi_{u,n}=1)\times p(\xi_{u,n}=1)} \\ &\quad\times \frac{\prod_{\{n:\xi_{u,n}=0\}} p(w_{u,n}|z_{u,n})\times p(z_{u,n}|z_{u,n-1},\xi_{u,n}=0)\times p(\xi_{u,n}=0)}{\prod_{\{n:\xi_{u,n}=0\}} p(w_{u,n}|z_{u,n})\times p(z_{u,n}|z_{u,n-1},\xi_{u,n}=0)\times p(\xi_{u,n}=0)} \\ &\quad= \frac{\prod_{t=1}^V \Gamma\left(\beta_t+n_t^{(z_{u,n-1})}\right)}{\prod_{t=1}^V \Gamma\left(\beta_t+n_{t,\neg(u,n)}^{(z_{u,n-1})}\right)} \times \frac{\Gamma\left[\sum_{t=1}^V\left(\beta_t+n_{t,\neg(u,n)}^{(z_{u,n-1})}\right)\right]}{\Gamma\left[\sum_{t=1}^V\left(\beta_t+n_t^{(z_{u,n-1})}\right)\right]} \times \frac{\mathrm{Beta}\left(\gamma_0+n_u^{(1)},\gamma_1\right)}{\mathrm{Beta}\left(\gamma_0+n_{u,\neg(u,n)}^{(1)},\gamma_1\right)} \\ &\quad\propto \frac{\beta_t+n_{w_{u,n},\neg(u,n)}^{(z_{u,n-1})}}{\sum_{t=1}^V\left(\beta_t+n_{t,\neg(u,n)}^{(z_{u,n-1})}\right)} \times \frac{\gamma_0+n_{u,\neg(u,n)}^{(1)}}{\gamma_0+\gamma_1+n_{u,\neg(u,n)}^{(0)}+n_{u,\neg(u,n)}^{(1)}} \qquad \text{if } k=z_{u,n}\end{aligned} \tag{S.11}$$

and

$$p(z_{u,n}=k,\xi_{u,n}=1|\mathbf{z}_{\neg(u,n)},\boldsymbol{\xi}_{\neg(u,n)},\mathbf{W})=0 \qquad \text{if } k\neq z_{u,n} \tag{S.12}$$

Similarly, we can obtain Equation (3),

$$\begin{aligned}&p(z_{u,n}=k,\xi_{u,n}=0|\mathbf{z}_{\neg(u,n)},\boldsymbol{\xi}_{\neg(u,n)},\mathbf{W}) \\ &\quad= \frac{\prod_{\{n:\xi_{u,n}=0\}} p(w_{u,n}|z_{u,n})\times p(z_{u,n}|z_{u,n-1},\xi_{u,n}=0)\times p(\xi_{u,n}=0)}{\prod_{\{n:\xi_{u,n}=0\}\neg(u,n)} p(w_{u,n}|z_{u,n})\times p(z_{u,n}|z_{u,n-1},\xi_{u,n}=1)\times p(\xi_{u,n}=0)}\end{aligned}$$

$$
\begin{aligned}
&\propto \frac{\frac{\Gamma\left(\sum_{t=1}^{V}\beta_t\right)}{\prod_{t=1}^{V}\Gamma(\beta_t)} \times \frac{\prod_{t=1}^{V}\Gamma\left(\beta_t+n_t^{(k)}\right)}{\Gamma\left[\sum_{t=1}^{V}\left(\beta_t+n_t^{(k)}\right)\right]} \times \frac{\Gamma\left(\sum_{k=1}^{K}\alpha_k\right)}{\prod_{k=1}^{K}\Gamma(\alpha_k)} \times \frac{\prod_{k=1}^{K}\Gamma\left(\alpha_k+n_u^{(k)}\right)}{\Gamma\left[\sum_{k=1}^{K}\left(\alpha_k+n_u^{(k)}\right)\right]} \times \frac{\mathrm{Beta}\left(\gamma_0,\gamma_1+n_u^{(0)}\right)}{\mathrm{Beta}(\gamma_0,\gamma_1)}}{\frac{\Gamma\left(\sum_{t=1}^{V}\beta_t\right)}{\prod_{t=1}^{V}\Gamma(\beta_t)} \times \frac{\prod_{t=1}^{V}\Gamma\left(\beta_t+n_{t,\neg(u,n)}^{(k)}\right)}{\Gamma\left[\sum_{t=1}^{V}\left(\beta_t+n_{t,\neg(u,n)}^{(k)}\right)\right]} \times \frac{\Gamma\left(\sum_{k=1}^{K}\alpha_k\right)}{\prod_{k=1}^{K}\Gamma(\alpha_k)} \times \frac{\prod_{k=1}^{K}\Gamma\left(\alpha_k+n_{u,\neg(u,n)}^{(k)}\right)}{\Gamma\left[\sum_{k=1}^{K}\left(\alpha_k+n_{u,\neg(u,n)}^{(k)}\right)\right]} \times \frac{\mathrm{Beta}\left(\gamma_0,\gamma_1+n_{u,\neg(u,n)}^{(0)}\right)}{\mathrm{Beta}(\gamma_0,\gamma_1)}} \\
&\propto \frac{\beta_{w_{u,n}}+n_{w_{u,n},\neg(u,n)}^{(k)}}{\sum_{t=1}^{V}\left(\beta_t+n_{t,\neg(u,n)}^{(k)}\right)} \times \frac{\alpha_k+n_{u,\neg(u,n)}^{(k)}}{\sum_{k=1}^{K}\left(\alpha_k+n_{u,\neg(u,n)}^{(k)}\right)} \times \frac{\gamma_1+n_{u,\neg(u,n)}^{(0)}}{\gamma_0+\gamma_1+n_{u,\neg(u,n)}^{(0)}+n_{u,\neg(u,n)}^{(1)}} \qquad \text{(S.13)}
\end{aligned}
$$

Appendix D: Summary of the Recommendation Methods Applied in Our Experiments

Table S.2 Description of the Benchmarks and the Proposed PE-LDA

Method	Category	Basic Idea
Pop-Rank	Non-personalized Method	This method represents a baseline model and recommends the top-N most popular followees to follow. Popularity is calculated by the number of followers in our experimental dataset.
Logistic-MF	MF	As one of the state-of-the-art MF extensions for implicit feedback data, this method uses a probabilistic framework to learn the probability that a user will be interested in a specific followee, and makes recommendations accordingly.
BPR-MF	MF	As one of the state-of-the-art MF extensions for implicit feedback data, this method uses a probabilistic framework and adopts the Bayesian optimization method for followee recommendations.
LDA	LDA-based	This is an implementation of the LDA method that leverages the repository of following lists to learn the user embeddings and make followee recommendations.
PE-LDA (This paper)	LDA-based	This is an implementation of our proposed PE-LDA model. Guided by the causal relationship between contextual engagement and decision making, this method combines the general preferences and contextual engagements in users' following sequences to learn the dynamic user representations and make followee recommendations.

Note. "MF" = Matrix Factorization, "BPR-MF" = Bayesian Personalized Ranking MF, "LDA" = Latent Dirichlet Allocation, "PE-LDA" = Preference-Engagement LDA.

Appendix E: Convergence Analysis

To analyze the computational cost required for the Gibbs sampler to generate stable results, we run the LDA-based methods on Twitter Friends dataset in this appendix for model fit analysis where Gibbs sampling will be iterated 500 times. Figure S.2 shows the in-sample fit of PE-LDA and LDA with $K = 10$, 30, and 50, respectively, where the log-likelihood of the models versus the iteration times are reported. The results are averaged over the 5-fold cross-validation. Figure S.2 reveals that, in terms of in-sample fit, the topic

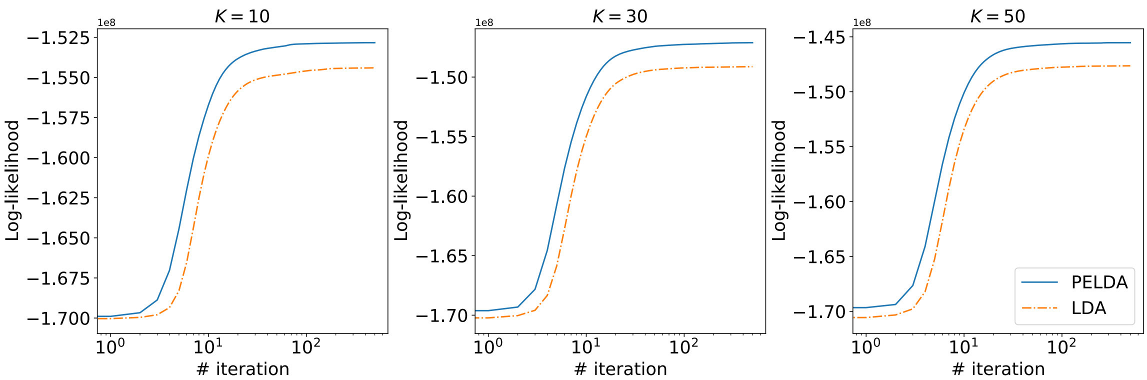
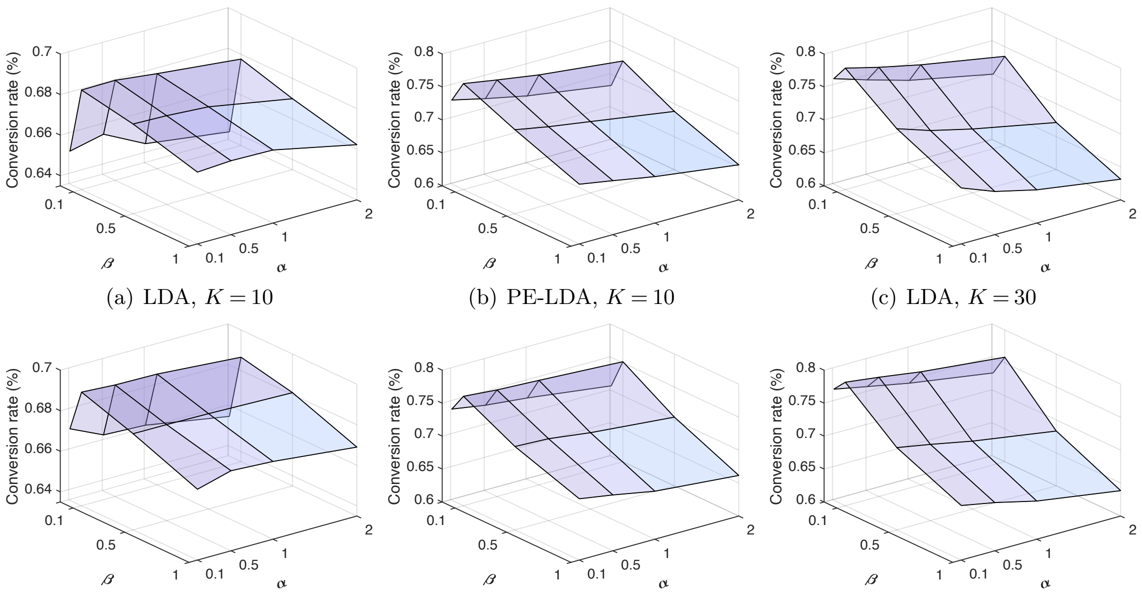


Figure S.2 Log-likelihood w.r.t. The Number of Iterations for LDA-based Methods

model with contextual engagement incorporated (i.e., PE-LDA) is preferred in general over the model with i.i.d. topic assignments (i.e., LDA). This is evidenced by the higher log-likelihood of PE-LDA. Figure S.2 also shows a fast burn-in period of PE-LDA (nearly stable after 50 iterations), indicating that PE-LDA can achieve stable results without extra iteration cost.

Appendix F: Sensitivity Analysis: Impacts of α and β

To further investigate the potential influence of the hyperparameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ on recommendation performance, we conduct parameter tuning for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ for LDA-based methods in this appendix. Since we do not have any prior knowledge about the distribution of each followee on each topic and that of each topic on each user, we set $\alpha_j = \alpha$ $(j = 1, ..., K)$ and $\beta_j = \beta$ $(j = 1, ..., V)$ and tune $\alpha \in \{0.1, 0.5, 1, 2\}$ and $\beta \in \{0.01, 0.1, 0.5, 1\}$ following previous research (e.g., Xu et al. 2018). As shown in Figure S.3, both methods



(a) LDA, $K = 10$ (b) PE-LDA, $K = 10$ (c) LDA, $K = 30$

(d) PE-LDA, $K = 30$ (e) LDA, $K = 50$ (f) PE-LDA, $K = 50$

Figure S.3 Joint Impact of Hyperparameters α and β on Recommendation Performance in Terms of Conversion Rate (All Results are Generated from Top-10 Recommendation)

reach the best performance at $\beta = 0.1$. This is consistent with the default setting that is widely applied in the related work. It can also be observed that the performance of PE-LDA and LDA is more sensitive to β (vs. α). Specifically, the conversion rate reaches its highest value at $\beta = 0.1$, and then degradates sharply as β continues to increase. This becomes increasingly evident when the number of interests (K) increases for both PE-LDA and LDA.

We provide a possible explanation for the above phenomenon as follows. Drawing upon the fact that high α and β priors tend to result in less decisive interest associations (i.e., the model prefers to assign more interests to each user with a higher α, and to assign more followees to each interest with a higher β; Blei et al. 2003), the above results indicate that the interest–followee distribution $\boldsymbol{\varphi}$ (of which the conjugate prior is Dir(β)) seems more influential than the user–interest distribution $\boldsymbol{\theta}$ (of which the conjugate prior is Dir(α)) on recommendation results. Since there are usually far more followees than interests on an OSN, the unit change of β would result in more (or less) items (most of which are interference items) in shaping the distribution (e.g., $\boldsymbol{\varphi}$) than that of α does. Thus, the unit change of β (vs. α) may cause more (or less) false positives in recommendation, which indicates a stronger sensibility of β. Therefore, to ensure the effectiveness of the LDA-based recommendation methods, high α and β (especially β) priors are usually not preferable.