

Unraveling Multifaceted User Preferences on Content Platforms: A Bayesian Deep Learning Approach

Mingzhang Yin*, Ziwei Cong†, Jia Liu‡

Abstract

Given the increasing importance of user engagement on digital content platforms, this paper proposes a Bayesian deep learning model, called the Multi-Dynamic Neural Poisson System (MDNPS), that can capture the multifaceted nature of user activities on such platforms. The MDNPS yields semantically interpretable factors shared across time and user activities over high-dimensional unstructured content, quantifies user preferences at different granularity levels, and adeptly captures their temporal and cross-activity dependencies. This model is scalable to large empirical data and can be inferred efficiently with a stochastic variational Bayes algorithm. We apply the MDNPS to the largest knowledge-sharing platform in China, focusing on the dynamics in user content consumption and contribution behaviors. We show that the MDNPS significantly outperforms benchmark factor models for both in-sample and out-sample data fitting. Our platform-level and individual-level estimates show rich and interpretable insights about consumer preference dynamics and their relationships with user popularity. These insights offer valuable managerial implications for platforms to design content strategies, enhance user engagement, and sustain platform growth.

Keywords: User-Generated Content, Preference Measurement, Digital Platforms, Dynamic Model, Neural Networks

*Mingzhang Yin is an Assistant Professor of Marketing, University of Florida. Email: mingzhang.yin@warrington.ufl.edu

†Ziwei Cong is an Assistant Professor of Marketing, Georgetown University. Email: zc260@georgetown.edu

‡Jia Liu is an Associate Professor of Marketing and an affiliated Associate Professor of IEDA, Hong Kong University of Science and Technology. Email: jialiu@ust.hk. The research is funded by the General Research Fund (GRF) of Hong Kong Research Grants Council.

1 Introduction

With the advent of digital technology, online platforms have emerged as primary venues for consumers to interact with content. For instance, 71% of U.S. consumers get news content via social media, 84% engage in online communities to share their experiences and hobbies, and 58% turn to the Internet to seek customized solutions (Horri gan, 2001; Fox, 2007). This digital transformation has unlocked rich consumer behavior data, enabling platforms to understand user preferences deeply and develop effective business strategies. For instance, platforms can use consumer preferences to provide recommendations, personalize search results, serve relevant ads, detect experts, and identify meaningful trends (Hemann and Burbary, 2013).

Despite the exponential growth in data collection on digital content platforms, extracting interpretable and meaningful insights about user preferences involves several challenges that extant models cannot address simultaneously. Unlike conventional e-commerce sites that offer a finite assortment of products, content platforms present an almost infinite and ever-increasing number of alternatives (Dhillon and Aral, 2021).¹ However, individual users engage with only a small subset of these contents, which requires efficient dimension-reduction techniques. More importantly, users' preferences on digital platforms are highly heterogeneous across multiple dimensions. The users' behavior is shaped by external influences (i.e., platform-wide social trends) and personal experiences; their preferences are activity-specific (Cain, 2022) and evolve over time (Dew et al., 2020). For example, a user actively *generating* science and technology content may be interested in *consuming* content about Bitcoin; the user's interactions with Bitcoin content are influenced not only by personal taste but also by short-term trends and long-term seasonality on the platform. Thus, a joint model capturing multifaceted user behaviors is crucial for preference measurement with consistent interpretation and temporal continuity. This approach can help platforms form a comprehensive understanding of user behavior patterns, ultimately enhancing user engagement and sustaining platform growth.

This paper addresses these challenges by introducing a Multi-Dynamic Neural Poisson System (MDNPS) to infer multifaceted user preferences from high-dimensional, dynamic time series tensors. MDNPS unravels the heterogeneity and interdependence of user preferences over time and across various user activities, at both platform and individual levels. Built on a novel Bayesian deep learning framework, it combines probabilistic models (e.g., Bayesian methods) with deep neural networks (Wang and Yeung, 2020). The in-

¹This can be attributed to two reasons. First, content platforms typically offer unlimited storage space for content creators. As a result, creators have little incentive to delete their old content (Sun and Zhu, 2013). Second, unlike standard products, content is highly differentiated, so no two contents are identical.

trinsic advantage of the MDNPS is its interpretability and flexibility. Traditional Bayesian methods, despite their transparent model structures and rigorous inference, often face a rigidity issue because it is challenging to specify proper prior and likelihood distributions manually for complex data (Dew et al., 2024). In comparison, deep neural networks excel at automatically modeling high-dimensional data and complex nonlinear patterns but often fail in interpretability due to their black-box structure (Goodfellow et al., 2016). MDNPS integrates deep learning with a probabilistic factor model to achieve interpretability, scalability, and flexibility. This paper is among the first in marketing to apply Bayesian deep learning for preference measurement.

The MDNPS has several distinctive features. First, it uses a global and local factor decomposition to disentangle individual preferences from platform-wide herd behavior. Specifically, individual-level preferences are modeled as heterogeneous offsets to the uniform platform activity strength. In this hierarchical structure, the individual-level temporal dynamics are modeled by a correlated hidden Markov model (HMM), and the platform-level temporal dynamics are modeled by the Long Short-Term Memory (LSTM) neural networks. Notably, the correlated HMM allows for interpretable interactions across different user activities, where the hidden state of one activity is influenced by the states of *all* activities at the previous time step. Meanwhile, LSTM neural networks capture complex short-term and long-term trends. A stochastic variational Bayes algorithm is developed to enable efficient and scalable Bayesian inference. These characteristics make the MDNPS suitable for various digital and offline settings that involve multiple large-scale, high-dimensional, and dynamic user activities.

We apply the MDNPS to the largest online knowledge-sharing platform in China. The platform offers a rich setting for studying multifaceted user preferences, as most users continuously engage in content contribution and consumption, which are simultaneously influenced by their personal experience and platform-level trends. Our study models the activities of thousands of active users from 2011 to 2017, focusing on answering questions (i.e., reflecting user contribution preferences) and subscribing questions (i.e., reflecting user consumption preferences). The MDNPS demonstrates superior performance in both in-sample and out-of-sample data fitting compared to various static and dynamic benchmark models.

Substantively, the MDNPS generates rich and digestible insights and managerial implications from large-scale, high-dimensional data. At the platform level, it summarizes sparse user-tag interactions into 20 interpretable latent topics and captures the dynamics of user interests across the topics. These outputs help platforms to identify asynchronies between content contribution and consumption, enabling proactive content supply adjustments. Additionally, we use these outputs to design metrics that characterize topic

attributes such as persistence, burstiness, and emergence time. Our results indicate that content creation trends typically drive consumption trends, emphasizing the need to incentivize expert contributions. At the individual level, the MDNPS uncovers the evolution of user preferences over time. We show that the evolution depends greatly on users’ initial interests and expertise, with significant heterogeneity across users. The model’s individual-level parameters offer insights into user segmentation, preference migration patterns, preference discrepancy between consumption and contribution, alignment between individual preferences and aggregate platform trends, and the relationships between preferences and user popularity (i.e., eventual follower size). These insights demonstrate the value of MDNPS in offering a rich and deep understanding of platform users, which ultimately guides platforms’ content and targeting strategies.

The rest of the paper is organized as follows. We first review relevant literature, followed by our research context. Next, we specify the MDNPS and its inference algorithm, followed by model validation. We present and discuss various substantive outputs of our application. Finally, we conclude with a discussion.

2 Related Literature

This paper contributes to the literature on modeling consumer preferences. The classic approach by [Guadagni and Little \(1983\)](#) models brand preferences using exponential smoothing of customer-level brand loyalty parameters. Subsequent studies have incorporated product interdependencies ([Duvvuri et al., 2007](#)) and interactions between consumers ([Yang and Allenby, 2003](#)). More recent work has focused on dynamic user preferences for physical products and a single kind of consumer activity ([Dew et al., 2020; Li and Ma, 2020; Dhillon and Aral, 2021; Jacobs et al., 2021](#)). [Liu and Dzyabura \(2021\)](#) assumes multiple partworth vectors associated with each individual, and models the purchase utility as the maximal utility generated by these attribute partworths. Our model advances this line of research by modeling preferences not on the observed products but on the latent factors. The multifaceted preference in our study is shaped by the activities users engage in and the timing of these activities, which lead to individual heterogeneity that are prevalent on digital platforms.

Our proposed method builds on the literature of factor models, particularly Poisson factor analysis (PFA), a model widely used for high-dimensional, sparse count data in marketing ([Toubia, 2021; Liu et al., 2021, 2024](#)). In particular, the Poisson–Gamma Dynamical Systems (PGDS) adopts a state transition matrix to capture the short-term evolution of latent factors ([Schein et al., 2016](#)), while subsequent works introduce

hierarchical models to account for long-term historical influences on topic intensities (Gong et al., 2017; Guo et al., 2018). Our MDNPS extends the PFA framework by introducing deep neural networks to jointly model long-term, short-term, and cross-activity interdependencies.

Our methodology contributes to the growing marketing applications of deep learning for dynamic modeling (Liu, 2023; Malhotra et al., 2023). For instance, Recurrent Neural Networks (RNNs) such as the LSTM (Yu et al., 2019) have been used for predictive tasks (Liu et al., 2019; Chakraborty et al., 2022). In comparison, we use the LSTM to improve the flexibility of the prior and posterior distributions in a Bayesian framework. Recent marketing studies have begun using probabilistic methods to infer deep models (Dew et al., 2022; Burnap et al., 2023), but the resulting embeddings often lack interpretability, relying on heuristic regularization (Sisodia et al., 2023) or post hoc feature engineering (Cheng et al., 2022). In contrast, MNDPS anchors the interpretability in its transparent factor model and Bayesian inference. Therefore, compared to purely black-box predictive models, MDNPS overcomes the sparsity problems and offers interpretable measurements of user preferences.

Finally, our application context is related to the literature on user interest dynamics on digital content platforms. Previous studies have primarily focused on dynamics driven by short-term external events or platform interventions (Dhillon and Aral, 2021; West et al., 2021; Yin et al., 2022; Liberali and Ferecatu, 2022; Liu and Cong, 2023). We complement this literature by considering users' long-run evolution throughout their lifetime on content platforms, which generates important managerial implications for dynamic targeting and segmentation. In the context of knowledge-sharing platforms, our work enriches the understanding of user online learning and self-improvement (Koksal, 2020; Goli et al., 2022). Unlike traditional education research that relies on cross-sectional data, we explore within-individual learning trajectories using granular data. Our model and empirical findings can be useful for many online knowledge-sharing platforms (e.g., Quora, Coursera, Reddit, edX, Udacity) to improve user learning experiences.

3 Empirical Context

3.1 Zhihu

Our application context Zhihu had approximately 100 million monthly active users, 44 million questions, and 2.4 billion answers by the end of 2020 (Ouyang, 2021). The key features of Zhihu are similar to those of Quora, a well-known American Q&A platform. Zhihu users can seek knowledge, expertise, experiences,

and customized solutions by posting questions. A typical question consists of a title and several tags that indicate the themes of the question. A question usually receives multiple answers ranked by the number of votes they have received. See Figure 1 for an example. In addition, users can subscribe to questions of interest to receive notifications of newly posted answers. Users can follow any other users, and each user's follower size is a strong indicator of her expertise and reputation in this online community.

Topic tags: NBA, NBA 球员, 斯蒂芬·库里 (Stephen Curry), 巨星

Question title: 如果把库里放到上个世纪 90 年代，还能否有巨星的表现?

Answer 1: 张佳玲 (2021 新始答主)
765 人赞同了该回答
巨星在那个时代都会是巨星，只看不同教练治下，能巨成什么样了。
像马克·杰克逊带库里三年，库里也就是全明星、联盟二阵。科尔一用库里，库里立刻带队进王朝。
也不奇怪。达拉斯版本的纳什就是全明星，太阳版本的纳什就是MVP。太阳版本的基德就是个杰出后卫，网版本的基德就是准MVP。
看配阵，看教练。
怎么用很重要。
展开阅读全文

Answer 2: 漫无三点 (2021-04-29 12:41)
12 人赞同了该回答
能。
上世纪90年代，非法防守还没被取消，只要球队不是傻缺，库里大概率能得到大量的一对一机会。而且就算不能包夹弱侧无球人这一点，以库里的无球跑位穿插能力，对手得头疼死。
至于handcheck，对喜欢阵地持球三威胁的人有用，对库里真没多大用
发布于 2021-04-29 12:41

Answer 3: Air真心为你 (动漫爱好者)
235 人赞同了该回答
曾经有位知名答主 @静易居 写过一个这样的回答如果巅峰奥尼尔现在的 NBA 打球会怎样？里面的论点本人非常赞同
他的论断是：即便奥尼尔放到现在的联盟，同样具有高度的战略价值
所以基于同样的理由，你问库里放到别的年代会怎样，我同样会说，库里放到别的年代，同样具有高度的战略价值！
首先，库里和奥尼尔，代表着篮球这项运动的两个极端，而且都是各自领域的佼佼者，甚至库里更bug。毕竟强如奥尼尔，你也依然能从漫漫历史长河中找到许多和他迥然不同的人，张伯伦，拉塞尔，贾巴尔，大梦，罗宾逊，尤因，他们虽然风格不尽相同，但都是具备超强统治力的内线人物。但是，库里在历史长河中，你找不到一个模板，没错，一个也没有。即使速如密阿伦，米勒这两位顶级射手，他们会的库里的也会，甚至比他们更会，库里会的，他们却未必会。
所以，库里和奥尼尔，代表着篮球这项运动的两个极端，就像答主 @静易居 说的一样：奥尼尔是历史开阅读全文

Figure 1: Example of a Question and its Answers.

The Q&As on Zhihu span thousands of different topics, covering all walks of life, including finance, healthcare, the Internet, information technology (IT), education, social issues, and entertainment. According to a 2017 report by iClick and Zhihu², approximately 70% of users leverage Zhihu for knowledge acquisition

²The original report (in Chinese) is currently unavailable online. However, some of its key findings and statistics (in English) can be found here: <https://www.dragonsocial.net/blog/zhihu-marketing-guide/>.

and self-improvement, 68% express a desire to subscribe and discuss topics of interest, and 43.1% wish to share their insights and experiences. Therefore, understanding users' topical interests/expertise and facilitating their learning/exploration journey is critical to the platform and user welfare. More generally, given the widespread popularity of Zhihu, its users' behavior and topical trends can provide valuable representations of the broader population, especially among the well-educated young Chinese-speaking demographics.

3.2 Data Description

We have rich data sets on the historical activities of all the registered users from its launch in January 2011 until June 2017. We focus on modeling user answer contributions and question subscriptions because they are the most frequently observed user activities and the most important indicators of user preferences.³ Specifically, answered questions reflect user domain expertise, while subscribed questions reflect user consumption preferences. We construct our sample using the following selection criteria. First, we focus on user activities from the first month to the 30th month since each user's registration. This criterion permits comparing and synchronizing the "personal clocks" across different users. Second, we confine the sample to users who joined the platforms before January 1, 2015, so we have at least 30-month observations per user. Third, to ensure enough data for understanding the dynamics in individual users' multifaceted preferences, we only include users with at least one answer contribution and one question subscription activity every three months since joining Zhihu.

These selection criteria resulted in a sample of 14,902 users, which offers a relatively representative view of the most active and important content contributors on Zhihu. By the end of the 30th month on Zhihu, an average sample user contributed 666 answers ($S.D. = 1,121$) and subscribed to 2,075 questions ($S.D. = 4,484$). These questions are related to 19,439 tags, reflecting broad and diverse interests among users. Despite our selection of relatively active users, we still find large activity sparsity within each user and high heterogeneity across users. See Figure W1 in Web Appendix A for details. These observations confirm the importance of using sophisticated models to understand user preferences effectively in this context.

³We do not model the activity of question contribution due to data sparsity (an average sample user only contributed 1.311 questions in an average month).

4 Multi-Dynamic Neural Poisson System

Our proposed MDNPS is a deep dynamic factor model that captures the evolution of user multifaceted preferences on digital content platforms, where users often dynamically interact with many items (e.g., tags, topics, products, and users) in multiple ways (e.g., answering and subscribing questions). In this section, we first describe the model specification using our empirical context, but it can be generalized to similar platforms. Then, we introduce an efficient algorithm for model inference.

4.1 Model Specification

Each question on Zhihu is labeled with a small set of tags. Let the corpus consist of N users, J tags, M types of user activities, and T periods. We model the time at the monthly level, a time window that balances activity richness and granularity. Both N and J are large on Zhihu, but most users interact with only a small subset of tags, resulting in sparse user-tag interactions. This large-scale yet sparse data structure complicates tag-level analysis and motivates us to adopt the factor model, particularly the Poisson factor analysis, for dimension reduction. We focus on $M = 2$ user activities: answering questions and subscribing to questions on Zhihu. Figure 2 provides the graphical model of the MDNPS.

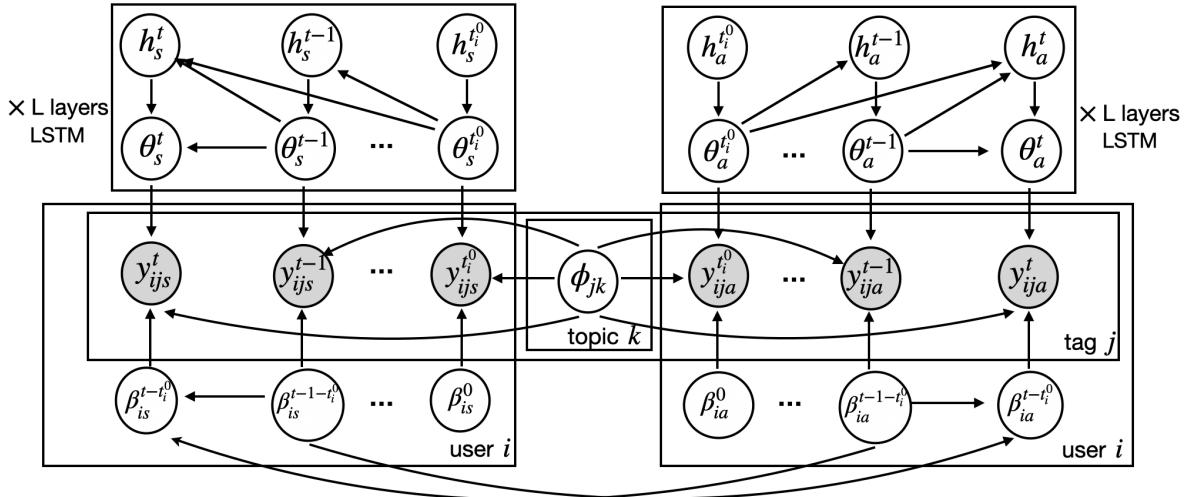


Figure 2: Graphical Model of the MDNPS.

Note: The subscripts ‘a’ and ‘s’ represent question answering and subscription, respectively. t_i^0 is the calendar time when user i joined the platform. The superscript is the user relative lifetime for β and is the calendar time for other variables.

4.1.1 Topics

Suppose K topics (a.k.a. factors) in this corpus are represented by the time-invariant $J \times K$ matrix ϕ . Each $J \times 1$ vector ϕ_k represents the tag distribution of topic k . Its prior distribution is specified on the logit space as a standard Normal distribution,

$$\tilde{\phi}_k \sim \mathcal{N}(0, I) \text{ and } \phi_k = \text{softmax}(\tilde{\phi}_k). \quad (1)$$

This prior distribution is equivalent to a standard logistic normal distribution for ϕ_k (Atchison and Shen, 1980). The softmax function $\text{softmax}(\phi) = \exp(\phi) / \sum_{j=1}^J \exp(\phi_j)$ makes ϕ_k a valid tag distribution for the topic k with $\sum_{j=1}^J \phi_{jk} = 1$. Thus, each topic can be interpreted based on the tags with the highest probabilities. Note that the same latent topics are shared across time periods and activities, which is key to maintaining a consistent definition of topics and for understanding how preferences evolve and interrelate across different topics.

4.1.2 The data likelihood

We let $y_{ijm}^t \in \mathbb{N}$ denote user i 's number of activity of type m for tag j at calendar time t , $i \in \{1, \dots, N\}$, $j \in \{1, \dots, J\}$, and $m \in \{1, \dots, M\}$. For each time t , the observation is a 3-way tensor $\mathbf{Y}^t = \{y_{ijm}\}_{i \in [N], j \in [J], m \in [M]}$, and the full data $\mathbf{Y} = (\mathbf{Y}^1, \mathbf{Y}^2, \dots, \mathbf{Y}^T)$ are high-dimensional dynamic tensors. We assume that y_{ijm}^t follows a Poisson distribution with a rate parameter $\mu_{ijm}^t > 0$,

$$y_{ijm}^t \sim \text{Pois}(\mu_{ijm}^t). \quad (2)$$

Users enter a digital platform at different periods, hence we do not observe the activities of user i for all the time $t \in \{1, \dots, T\}$. Thus, we let t_i^0 denote the time that user i joined the platform, and set $t_i = t - t_i^0$ as user i 's lifetime on the platform. We then define user-specific parameters based on user lifetime rather than calendar time. The lifetime t_i allows for individual-specific lifecycle on the platform, despite receiving the same platform-level shocks at a given time t .

Using the above-defined topics, we decompose the Poisson rate parameter as a weighted sum of the tag

distributions ϕ_{jk} across the K topics,

$$\mu_{ijm}^t = \sum_{k=1}^K \theta_{mk}^t \beta_{imk}^{t_i} \phi_{jk}. \quad (3)$$

The term $\theta_{mk}^t \beta_{imk}^{t_i}$ reflects user i 's total activity strength of a topic k at time t . Specifically, the parameter θ_{mk}^t captures the population preferences in topic k at time t , often influenced by the external major events and the platform trend. The parameter $\beta_{imk}^{t_i}$ captures user i 's preferences at lifetime t_i , reflecting how the user's interests evolve along with the aggregate population trends (i.e., similar to individual random effects). We use multiplication between θ_{mk}^t and $\beta_{imk}^{t_i}$ due to following considerations: (1) the rate of Poisson distribution being positive necessitates positivity of $\theta_{mk}^t, \beta_{imk}^{t_i}, \phi_{jk}$, and (2) if the interaction is defined additively as $\theta_{mk}^t + \beta_{imk}^{t_i}$, the individual preference can only amplify the population trends because $\theta_{mk}^t + \beta_{imk}^{t_i} \geq \theta_{mk}^t$; this would make the estimation of population trends inaccurate since the activity strength of an individual is not always stronger than that of the population. Instead, for our Poisson rate model in Equation (3), a $\beta_{imk}^{t_i}$ larger or lower than one can indicate whether user i reacts stronger or weaker than the trends related to topic k , hence θ_{mk}^t will encode the platform trend.

The factor model disentangles different factors driving users' topical preferences over time and allows for flexible dynamic modeling. The hierarchical structure enables the model to capture complex temporal dependencies in the evolution of platform trends θ_{mk}^t , while capture cross-activity dependence in the evolution of individual tastes $\beta_{imk}^{t_i}$. This model design balances flexibility with computational efficiency.

4.1.3 Platform-level dynamics and long-term dependency

We now describe the evolution of platform-level topic strength θ_{mk}^t . In general, the dynamic patterns of topics at the platform level are often complex. For example, an election season may drastically increase user interest in political content for a few months, which may also impact the popularity of political content in future election cycles, resulting in a long-term dependency. Furthermore, the popularity of a topic like "E-commerce" may depend on the dynamics of related topics (e.g., "Digital Media" and "Online Communities"), and this dependency is generally nonlinear. Given the complex cross-time and cross-topic dependence, we leverage the LSTM neural network to model the evolution of topic strength θ flexibly.

For ease of notation, we omit the subscript m in θ_{mk}^t when describing the LSTM specification, as the structure and process are the same across activity types. We use $\theta^t \in \mathbb{R}^K$ to denote the vector $(\theta_1^t, \dots, \theta_K^t)^\top$.

We first remove the positivity restriction by projecting each θ^t to the logarithmic space using an embedding matrix E as $\hat{\theta}^t = E \log(\theta^t)$. Then, we transform the sequence of embeddings $\{\hat{\theta}^0, \hat{\theta}^1, \dots, \hat{\theta}^{T-1}\}$ to a sequence of hidden states $\{h^1, h^2, \dots, h^T\}$ by the LSTM $f(\cdot)$,

$$(h^1, h^2, \dots, h^T) = f(\hat{\theta}^0, \hat{\theta}^1, \dots, \hat{\theta}^{T-1}), \quad (4)$$

where $\hat{\theta}^0 \sim \mathcal{N}(0, \mathbf{I}_K)$ is a padding vector to ensure that the hidden state $h^t(\hat{\theta}^{0:t-1})$ at time t only depends on the information $\hat{\theta}^{0:t-1}$ before t . The LSTM uses a forget gate f_t , an input gate i_t , and an output gate o_t to store important information and forget irrelevant information as follows:

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h^{t-1}, \hat{\theta}^{t-1}] + b_f), \quad i_t = \sigma(W_i \cdot [h^{t-1}, \hat{\theta}^{t-1}] + b_i) \\ \tilde{C}_t &= \tanh(W_C \cdot [h^{t-1}, \hat{\theta}^{t-1}] + b_C), \quad C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \\ o_t &= \sigma(W_o \cdot [h^{t-1}, \hat{\theta}^{t-1}] + b_o), \quad h^t = o_t \odot \tanh(C_t), \end{aligned} \quad (5)$$

where \tilde{C}_t is the cell gate, C_t is the cell state, and σ is the sigmoid function $\sigma(z) = 1/(1 + \exp(-z))$. The training parameters of the LSTM are the weight matrices W_f, W_i, W_C, W_o and bias vectors b_f, b_i, b_C, b_o .

With the hidden states $\{h^1, h^2, \dots, h^T\}$, the information before time t propagates to θ^t by

$$p(\theta^t | \theta^{1:t-1}) = \mathcal{LN}(\mu(\hat{\theta}^{t-1}, h^t(\hat{\theta}^{0:t-1})), \text{diag}(\sigma^2(\hat{\theta}^{t-1}, h^t(\hat{\theta}^{0:t-1}))))). \quad (6)$$

Here, $\mathcal{LN}(\mu, \text{diag}(\sigma^2))$ refers to the log-normal distribution that ensures the positivity of θ^t . $\text{diag}(\sigma^2)$ denotes a diagonal covariance matrix with σ^2 on the diagonal. $\mu(\cdot)$ and $\sigma(\cdot)$ are two feed-forward neural networks, which take the concatenated $(\hat{\theta}^{t-1}, h^t)$ as the inputs, and output the mean and the variance of the conditional distribution of θ^t , respectively. The key in Equations (4) and (6) is that the platform-level activity strength at time t associates with the information from all the previous time steps, thus capturing long-term dependence. We let W_θ denote all the parameters involved in this part, including the embedding matrix E , the neural network parameters $\{W_f, W_i, W_C, W_o, b_f, b_i, b_C, b_o\}$ of the LSTM $f(\cdot)$, and the parameters $\theta_\mu, \theta_\sigma$ of the feedforward neural networks $\mu(\cdot)$ and $\sigma^2(\cdot)$.⁴

⁴In practice, the LSTM in Equation (4) can be stacked to form an L -layer deep LSTM as $(h_1, h_2, \dots, h_T) = f_L f_{L-1} \cdots f_1(\hat{\theta}^0, \hat{\theta}^1, \dots, \hat{\theta}^{T-1})$ to improve flexibility. We adopt $L = 3$ in our empirical study.

4.1.4 Individual-level dynamics and cross-activity dependency

We now model the evolution of individual-level offset β_{imk}^t . Since the number of individual parameters is large and the individual-level data is sparse, we use a transparent hidden Markov approach instead of the LSTM to model β_{imk}^t for computational efficiency and accuracy. We assume that the K -dimensional vector $\beta_{im}^t := (\beta_{im1}^t, \dots, \beta_{imK}^t)^\top$ is influenced by user i 's preferences for all the activity types at the previous time point $t - 1$. In other words, we allow individual users' consumption and contribution preferences to influence each other over time. To this end, we specify β_{im}^t using a correlated Markov model as

$$p(\beta_{im}^t | \beta_{i,1:M}^{t-1}) = \mathcal{LN}\left(\sum_{m'=1}^M \Pi^{m'm} \tilde{\beta}_{im'}^{t-1}, \sigma_0^2 I\right). \quad (7)$$

where $\tilde{\beta}_{im}^{ti}$ denotes the log-transformed β_{im}^{ti} (to remove the positive constraints) and $\Pi^{m'm}$ are the transition matrices. Note that the topic intensity of activity m at time t depends on the topic intensities of all the activity types $m' = 1, \dots, M$ at time $t - 1$. By doing so, this correlated Markov model quantifies within and across activity-type transitions of individual preferences. It captures how a user's contribution (consumption) preferences in one topic transit to a contribution (consumption) preferences in the same and other topics.

4.2 Inference and Orthogonal Topic Regularization

We let $\eta := (W_\theta, \{\Pi^{m'm}\}, \sigma_0)$ denote all the parameters related to the specifications of θ and β , where $W_\theta = (E, W_f, W_i, W_C, W_o, b_f, b_i, b_C, b_o, \theta_\mu, \theta_\sigma)$. Given the above data likelihood and prior specifications, the joint distribution of the data and model parameters is

$$p(y, \theta, \beta, \phi; \eta) = \prod_{i,j,m} \left\{ \prod_t p(y_{ijm}^t | \theta_m^t, \beta_{im}^t, \phi_j) p(\beta_{im}^t | \beta_{i,1:m}^{t-1}; \sigma_0, \Pi) p(\theta_m^t | \theta_m^{1:t-1}; W_\theta) \right\} p(\phi_j). \quad (8)$$

Due to the high-dimensional neural network parameters, we develop a variational expectation-maximization (EM) algorithm (Wainwright et al., 2008) to jointly estimate the posterior distribution $p(\theta, \beta, \phi | y; \eta)$ (i.e. the inference step) and optimize the parameters η (i.e. the learning step). The intractable posterior is approximated by a set of variational distributions parameterized by the variational parameters ψ . For the variational distributions, we use the LSTM to model the joint posterior distribution of $\theta^{1:T}$, and use the mean-field distributions for β and ϕ ; see Web Appendix B for details. This choice of variational distribution maintains the long-term temporal dependency of θ and achieves high efficiency. The posterior inference is

conducted by finding the optimal variational parameter ψ that minimizes the Kullback-Leibler divergence between the variational distribution and the posterior, which is equivalent to maximizing tractable evidence lower bound (ELBO) objective (Blei et al., 2017).

With the proposed neural network architecture and the variational distributions, the inference can be conducted efficiently using the reparametrization trick, which computes the stochastic gradient of an unbiased ELBO estimation through backpropagation (Kingma and Welling, 2013). The proposed variational EM algorithm uses the adaptive moment method Adam (Kingma and Ba, 2014) to iteratively update the model parameters η and the variational parameters ψ until convergence. This update scheme corresponds to iteratively performing model fitting and posterior inference. We use the stochastic variational inference (Hoffman et al., 2013; Yin and Zhou, 2018) to update the parameters with minibatch data. The minibatch training allows fast parameter updates and efficient memory usage. Web Appendix B contains a complete specification of the modeling and inference structure and the variational EM algorithm.

Our optimization problem may be affected by topic collapse, a common issue observed in neural topic models (Li et al., 2022; Wu et al., 2023). Specifically, we find that the estimated platform-level tag interactions might distribute similarly for two different time points. This indicates that the algorithm converges to a locally optimal solution. Existing literature addresses the topic collapse issue using orthogonal regularization (Yao et al., 2014). However, this method cannot be directly applied to our problem, as it is usually designed for non-dynamic data and aims to reduce the similarity between factor loadings ϕ_k . Thus, we propose a new technique called orthogonal temporal regularization (OTR), which encourages orthogonality and distinctiveness across time t . The OTR involves a hyperparameter $\alpha \in [0, 1]$ that controls the regularization strength; we will choose α based on the topic quality defined in the following section. The details of the proposed OTR are provided in Web Appendix B.

5 Empirical Application

Our application has $N = 14,902$ users, $J = 19,439$ tags, $T = 72$ calendar months, 30-month user lifetime, and $M = 2$ activity types. We let $m = s$ indicate the content consumption behavior (i.e., subscribing to questions) and $m = a$ indicate the content contribution behavior (i.e., answering questions). We use the Pytorch as the auto-differentiation tool to compute the stochastic gradient of the ELBO. For each round of parameter update, we use a minibatch of 200 users to compute the objective of the stochastic variational

inference. We set 3 layers of 200 hidden units as our neural network architecture. We apply a small weight decay of 1.2×10^6 on all neural network parameters, following the convention in training deep models.

5.1 Topic Quality and Model Hyperparameters

We use the quality of extracted topics to determine the model hyperparameters (i.e., the number of topics K and the regularization strength α) because topic quality is critical for deriving meaningful insights from large-scale high-dimensional data. We measure topic quality by the product of topic coherence and topic diversity, following the existing literature (Dieng et al., 2019, 2020). Topic coherence quantifies the semantic similarity between high-probability words within a topic, indicating how meaningfully they are related to form a cohesive topic (Dieng et al., 2020). The coherence of a topic is high if its top tags tend to be associated with the same content. However, topics that collapse into a single topic have the highest coherence but the lowest quality. This concern is alleviated by considering topic diversity, which is defined as the proportion of distinct words among the top 25 words across all topics (Dieng et al., 2020).

We conduct a grid search for $K \in \{10, 15, 20, 25, 30\}$ and $\alpha \in \{0.05, 0.1, 0.3, 0.5\}$. Figure 3 shows the topic coherence and diversity for varying K with $\alpha = 0.1$ and varying α with $K = 20$. Overall, topic diversity decreases with K because the high-frequency tags are more likely to appear in different topics when K increases. The coherence peaks at $K = 20$ and then decreases. We also notice the topic diversity increases with α , which is consistent with our motivation in designing the OTR regularization. The coherence decreases as the diversity increases because the model is forced to incorporate irrelevant tags into the topics. Among the combinations, we choose $K = 20$ and $\alpha = 0.1$ as they give the highest topic coherence and most interpretable topics while keeping a relatively high diversity.

5.2 Benchmark Models

To understand the performance of the MDNPS in fitting the dataset, we consider potential alternative models for benchmark comparison. A naive approach is to model the data of each activity type and time slice with an independent topic model. However, this approach cannot generate consistent topics (as shown in the next subsection) and reduces data efficiency because only a small fraction of users interacted with a small subset of tags at each period. Another simplification would be a two-stage model that first learns a topic model and then uses a sequential model to fit the topic dynamics over time and across activities. However,

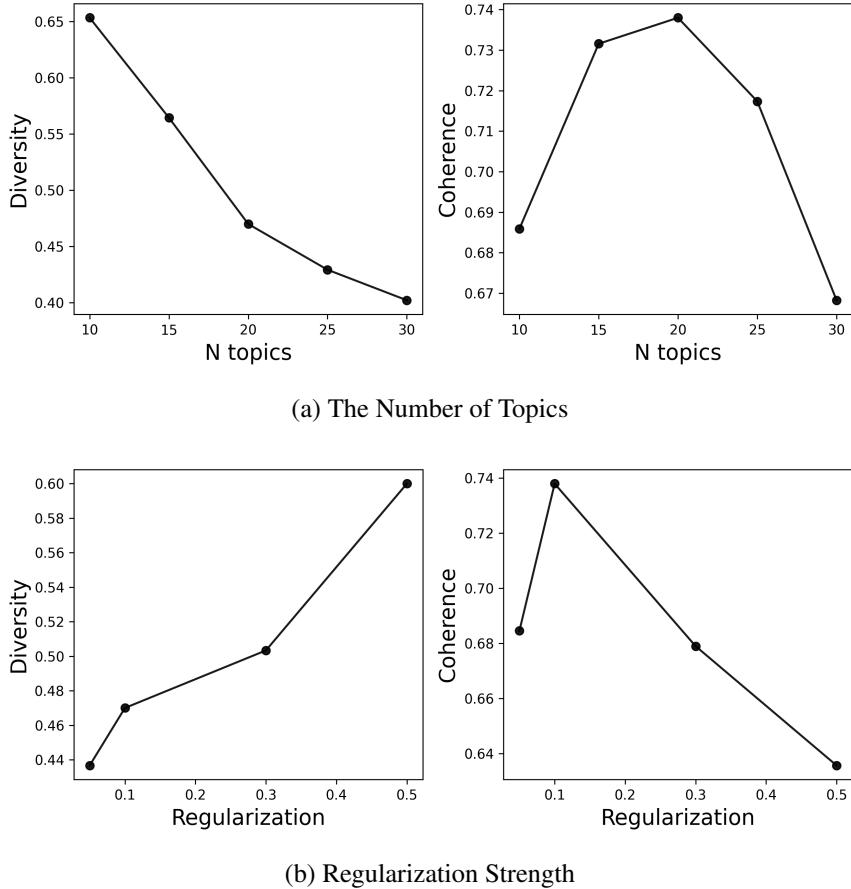


Figure 3: Selection of Model Hyperparameters.

this approach will result in discontinuities and inconsistencies in topic intensity evolution because it does not use information from adjacent time steps or related activities to inform the topic intensities. The advantages of integrating the learning of topics and their preferences evolution into a single, coherent framework have been theoretically and empirically validated in the dynamic topic model literature (Blei and Lafferty, 2006).

Given the above discussion, we compare the MDNPS against three benchmark factor models that can be applied to our data context. The first is the Bayesian Poisson tensor factorization (BPTF) introduced in Schein et al. (2015). The BPTF is a non-dynamic factor model. It factorizes a tensor in each of its order by assuming that for each activity type m , the observed count $y_{ijm} \sim \text{Pois}(\sum_{k=1}^K \theta_{imk}^{(1)} \theta_{jmk}^{(2)} \theta_{tmk}^{(3)})$. Each latent variable has an independent prior, so the BPTF does not model temporal or cross-activity dependence. Comparison with the BPTF can reveal the importance of joint modeling.

The second benchmark is the Deep Poisson-Gamma Dynamic System (Deep-PGDS) introduced in Guo

et al. (2018). It is a state-of-the-art dynamic model for count tensors, which uses recurrent neural networks to model temporal dependence. The Deep-PGDS differs from our MDNPS in two main aspects. First, it models the various user activities independently. Second, it only infers user preferences at the aggregated level. Comparison with the Deep-PGDS explores the benefits of jointly modeling multiple activity types and modeling individual-level preferences.

The last benchmark is a modified version of the MDNPS, denoted by the MDNPS\OTR, in which the OTR regularization is removed from the ELBO objective. This benchmark acts as an ablation study to illustrate the effectiveness of our proposed regularization technique in addressing the topic collapse problem. The hyperparameters in these benchmarks are similarly optimized using the topic quality metrics.

5.3 Comparison Results

We first compare these models on topic coherence, diversity, and quality in Table 1. The MDNPS performs significantly better than any benchmark model along all metrics ($p < 0.05$). The much higher coherence of the MDNPS relative to these benchmarks suggests that the MDNPS can better cluster tags with similar semantics, enabling easier and more precise interpretation. For example, the “Relationships” topic has top words as “Appearance, Intimacy, Childhood, Self-Management, Dating, Height, Goddess”, and the “Engineering and Sciences” topic has top words as “Agriculture, Civil Engineering, Manufacturing, Computer Major, China Stock Market, Cancer, C/C++.” The much higher diversity of the MDNPS relative to these benchmarks indicates its superiority in flexibly identifying distinctive topics. For example, our extracted topics cover a wide spectrum of Zhihu’s content ecosystems, including E-Commerce, Entertainment, Relationships, Politics, Mobile Device, Engineering and Sciences. Finally, we confirm that including OTR regularization significantly improves topic diversity (by addressing the topic collapse problem) and topic coherence (by grouping semantically related words to the same topics).

Next, we compare these models on fitting out-sample user activities in the 31st month (which is not used to train each model). We use standard metrics for evaluating factor models (Blei et al., 2003; Schein et al., 2019): the log-likelihood of the predictive distribution and the perplexity score (Wallach et al., 2009). Perplexity is computed as the exponentiated and reweighted likelihood of the hold-out data (i.e., a lower value indicates a better generalization from training data to new data). Thus, both metrics are likelihood-based, incorporating point prediction accuracy and uncertainty. The results are shown in Table 2. The MDNPS performs significantly better than all benchmarks along both metrics ($p < 0.05$), i.e., improving

Table 1: Topic Quality across Models.

Metric	BPTF	Deep-PGDS	MDNPS\OTR	MDNPS
Subscribing questions				
Coherence	0.695	0.474	0.716	0.751
Diversity	0.237	0.353	0.297	0.657
Quality	0.164	0.168	0.212	0.493
Answering questions				
Coherence	1.104	1.229	1.082	1.500
Diversity	0.287	0.460	0.347	0.657
Quality	0.316	0.565	0.375	0.985

Table 2: Out-Sample Predictive Performance across Models.

Metric	BPTF	Deep-PGDS	MDNPS\OTR	MDNPS
Subscribing questions				
Perplexity	1892.80	1535.90	1369.10	579.80
Log-Likelihood	-7.87	-7.67	-7.71	-6.92
Answering questions				
Perplexity	999.60	463.20	373.60	183.20
Log-Likelihood	-7.30	-6.49	-6.52	-5.79

Note: lower perplexity (higher log-likelihood) indicates a better fit.

perplexity by over 50% and increasing model likelihood close to 1 at the logarithmic scale for both activity types. Therefore, we conclude that the MDNPS can fit and predict the data better than these benchmarks.

6 Substantive Outputs

This section reports the estimated posterior mean of model parameters, including the latent topics $\{\phi_{jk}\}$, the platform-level topic dynamics $\{\theta_{mk}^t\}$, and the user-level preference trajectories $\{\beta_{imk}^{t_i}\}$. We showcase the validity of the MDNPS in reducing complex high-dimensional data into low and meaningful spaces, and in providing meaningful substantive insights.



Figure 4: Word Clouds of the Extracted Topics.

6.1 Topics

Our model depicts the content structure on the platform. The extracted topics summarize the sparse interaction between a large number of users and tags, revealing interpretable content structure at the topic level. We interpret each extracted topic k based on the tags with the highest probability weights in ϕ_k . Figure 4 presents the word clouds of the 20 extracted latent topics using the top 10 tags per topic.⁵ Aligned with the topic quality metrics in the previous subsection, we find the top tags of each topic are semantically similar and highly coherent. All topics are easy to interpret, except one that comprises a variety of categories, labeled as ‘‘Others.’’ These latent topics loosely correspond to three topical categories. The first category relates to IT and engineering, including ‘‘Mobile Device’’, ‘‘Tech-Finance’’, and ‘‘Engineering and

⁵The list of the top 20 tags per topic are shown in Table W1 of Web Appendix C.

Sciences.” The second category corresponds to social sciences such as “Culture & Religion,” “History & Novels,” and “Politics.” The third category is related to entertainment and daily life, including “Lifestyle,” “Relationships,” and “Entertainment and Trends.” Overall, these topics represent the information needs, expertise, and leisure interests of the Zhihu user community well.

6.2 Platform-Level Topic Trends

6.2.1 The evolution of topic popularity

The platform can understand the evolution of collective user behaviors by tracking $\{\theta_{mk}^t\}$ across topics over time, which reflects broad societal trends. To this end, we compute the normalized popularity per topic for each activity type in each calendar month, i.e., $\theta_{mk}^t / \sum_{t'=1}^T \theta_{mk}^{t'}$. This normalization does not change the dynamic pattern within a topic.

The left-side panel of Figure 5 reflects the platform-level content contribution trends. We find that in the early stage from January 2011 to March 2013, when Zhihu was a by-invitation-only online Q&A community, popular contributed topics were related to IT. Examples are “Mobile Devices”, “Zhihu Platform”, “Online Communities”, and “Social Network App.” This indicates that early adopters were likely tech enthusiasts or professionals. After the open registration in 2013, there was a noticeable diversification in topics. Contributions in “Sciences & Engineering”, “History and Novels”, and “Relationships” increased, indicating that the platform attracted a broader, more diverse audience after becoming publicly accessible. In addition, general and mainstream topics like “Entertainment and Trends” started gaining traction gradually.

The right panel of Figure 5 shows platform-level content consumption trends. Similar to contributions, consumption was initially dominated by IT-related topics, which decreased significantly after 2013. In addition, we find users’ consumption interests were timely associated with external major events. For instance, interests in “Tech Financ” and “Politics” are more responsive to external events, such as the rise of cryptocurrency and the U.S. presidential election. However, these interests were often short-lived, indicating transient herding behavior. Generally, the evolution of consumption interests significantly differs from but closely relates to contribution interests, further supporting our joint modeling of multifaceted user preferences.

These findings offer important managerial implications for digital content platforms. First, platforms should recognize the distinct preferences of early adopters and the general public. Their onboarding strategies should be tailored to the evolving interests and user base’s demographics. Second, the user interest

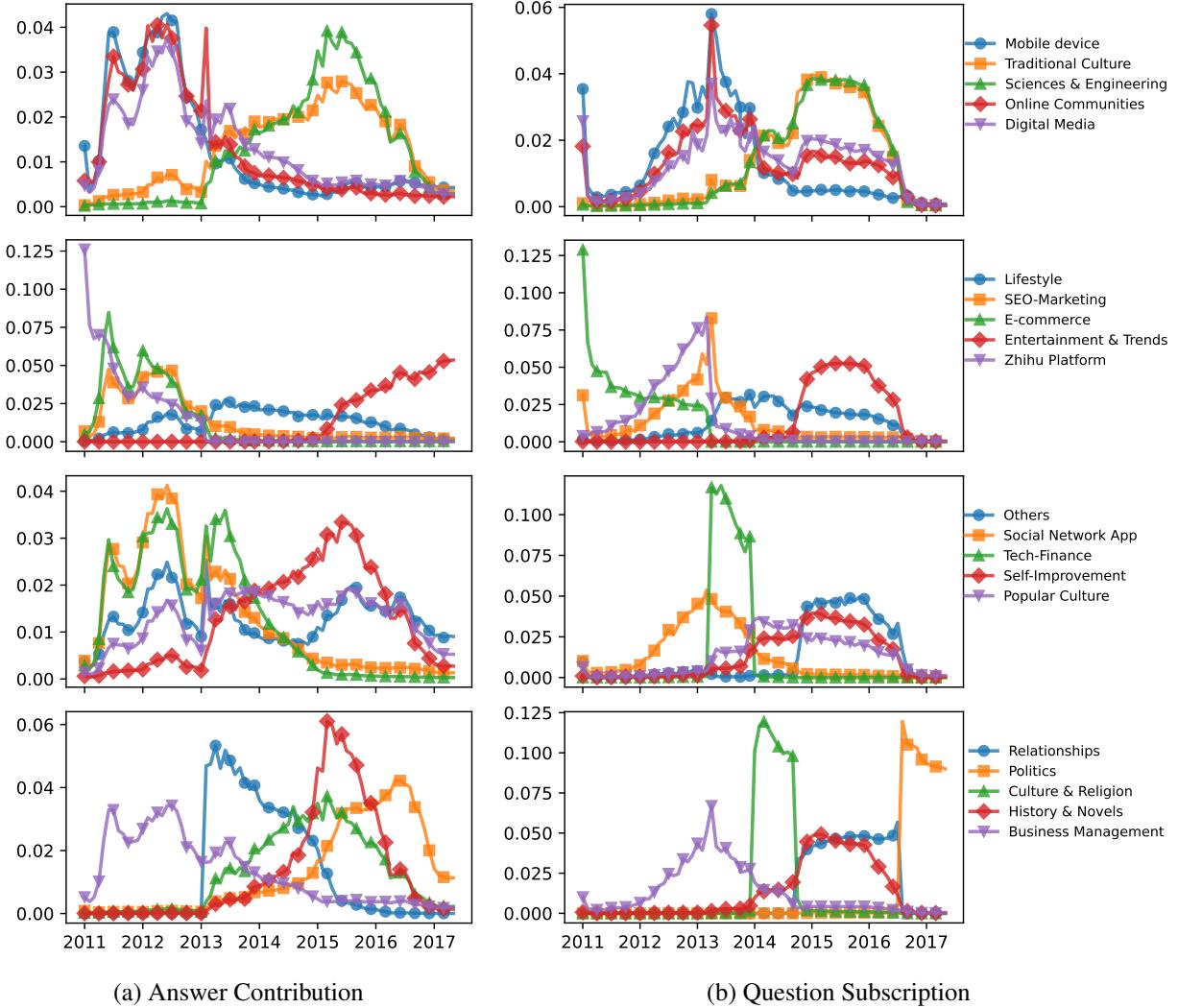


Figure 5: Platform-Level Within-Topic Dynamics by Activity Type.

spikes in response to external events underscore the need for timely content adaptation. Platforms should implement real-time monitoring tools that track consumption and contribution patterns and adjust their content strategies accordingly. This approach ensures that emerging trends and sudden interest spikes are promptly addressed. Third, while external events drive short-term interest, platforms should also focus on sustaining long-term engagement by fostering communities around enduring topics to maintain a steady user base. Similarly, platforms should diversify their content portfolio to include both niche and broad-interest topics. This can help build a more inclusive community, reflecting varied interests and fostering long-term engagement (compared with simply relying on transient spikes in user engagement).

6.2.2 Identifying gaps in content supply

Figure 5 suggests significant differences between content supply and demand. For example, the contribution to “Tech finance” was popular from 2011 to 2015, whereas its subscriptions became popular only in 2013. We introduce a ratio metric to understand the general gaps between content supply and demand. We first obtain the total intensity across time per topic k and activity type m as $\sum_{t=1}^T \theta_{mk}^t$. Then, we compute the ratio for topic k , denoted by Δ_k , as the total consumption intensity divided by the total contribution intensity. Thus, a higher ratio indicates a larger insufficient content supply relative to the content demand for topic k .

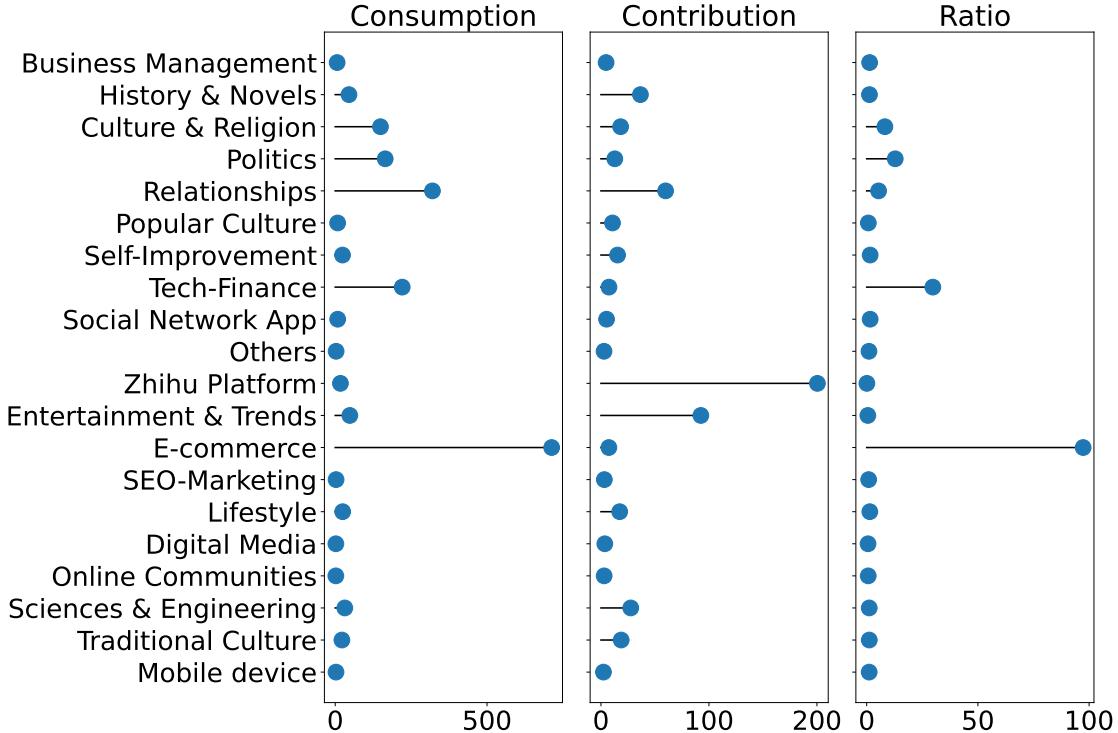


Figure 6: Aggregated Topic Strength.

For each topic, Figure 6 reports the total consumption intensity, contribution intensity, and their ratios. We find that topics with the highest ratios are “E-commerce”, “Tech Finance”, “Politics”, and “Culture & Religion.” It suggests that users consume more content than is being produced in these topics, indicating potential areas with unmet demand on which content contributors could focus their efforts. Platforms can incentivize experts and frequent contributors to focus on these areas through recommendations, rewards, or recognition programs.⁶

⁶For interested readers, we illustrate how our model outputs can be used for expert detection in Web Appendix D.

6.2.3 Topic persistence, emerge time, and burstiness

We introduce three additional topic characteristics: topic persistence, emerge time, and burstiness. Then, we highlight their key insights about topic dynamics across activity types.

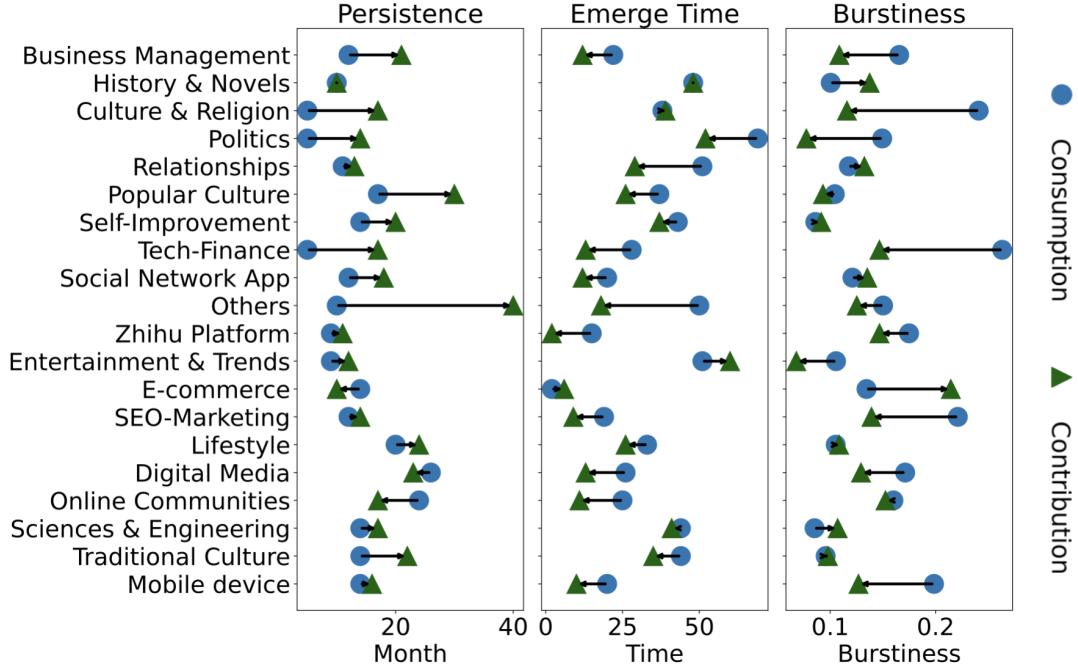


Figure 7: Topical Persistence, Emerge Time, and Burstiness.

Note: In the first panel, the x-axis represents the duration (in months) for which a topic remains popular on the platform. In the second panel, the x-axis indicates the specific calendar month when a topic began to gain popularity. In the third panel, the x-axis measures the intensity of burstiness.

Persistence. Persistence measures how long a topic remains relevant on the platform. Higher persistence indicates sustained user interests and continuous content contributions over time. Thus, identifying and maintaining persistent topics can enhance user retention and satisfaction (Shin et al., 2014), ultimately building a loyal user base for long-term engagement. We compute the cumulative strength of each topic k up to time T' as $C_{mk}(T') = \sum_{t=1}^{T'} \theta_k^t / \sum_{t=1}^T \theta_k^t$. We define the persistence of a topic as the time range between the lower and upper quartiles of $C_{mk} = (C(1), \dots, C(T_{mk}))$, i.e., $Q_{0.75}(C_{mk}) - Q_{0.25}(C_{mk})$ with $Q_\alpha(\cdot)$ as the α -quantile function. Thus, persistence is low if the strength of a topic concentrates around a few months and is otherwise high if the topic strength uniformly spreads over time.

The first panel of Figure 7 shows the persistence levels of each topic for consumption and contribution, respectively. On average, topics exhibit longer persistence in content contribution (18.3 months) than con-

sumption (12.9 months). The contribution persistence is longer than the consumption persistence for 17 of the 20 topics. These patterns suggest that while users' interest in consuming content may wane relatively quickly, contributors, particularly experts, remain engaged in their domains for extended periods (possibly due to the extensive time and effort required to develop expertise in new fields). Interestingly, topics that relate to everyday life rather than domain expertise, such as "Digital Media," "Online Communities," "Lifestyle," "Popular Culture," and "Self-Improvement," show high persistence in both consumption and contribution, indicating their ongoing relevance and importance in maintaining user engagement.

Emerge Time. Emerge time captures when a topic k starts gaining popularity. We define it as when the cumulative strength $C_{mk}(t)$ of topic k reaches its first quantile. This simple metric can provide insights into the diffusion of interests across topics. For example, it helps understand whether content consumption motivates content contribution or vice versa. The second panel of Figure 7 presents the consumption and contribution emerge time for each topic, separately. Overall, contribution interests generally emerge earlier than consumption interests for most topics (17 out of 20), with an average lead time of 9.3 months. This reveals that expert contributions often precede and likely drive user consumption. An exception is the "Entertainment and Trends" topic, where user interest peaks before contributions, which could be due to the immediate nature of news and trends. In sum, our findings suggest that platforms should focus on stimulating early contributions from experts to foster user interest and suggest what topics to stimulate. Encouraging expert participation can create a foundation for user-driven content consumption, particularly for topics where user interest follows experts' engagement.

Burstiness. Burstiness measures the intensity of fluctuations in topic strength over time. Understanding burstiness can help platforms manage real-time content strategies and adapt to changes in user interests. Highly bursty topics may require agile content management to capitalize on short-lived spikes in interest, while less bursty ones may benefit from steady content curation and promotion. Following Schein et al. (2019), we define the burstiness of a topic as the interest change at time t relative to its overall strength, i.e., $B_{mk} = \frac{1}{T-1} \sum_{t=1}^{T-1} |\theta_{mk}^{t+1} - \theta_{mk}^t| / \hat{\mu}_{mk}$, $\hat{\mu}_{mk} = \sum_{t=1}^T \theta_{mk}^t$.

The third panel of Figure 7 presents the consumption and contribution burstiness per topic. We find that topics such as "Tech-Finance," "Zhihu Platform," and "SEO-Marketing" exhibit high burstiness, indicating rapid changes in user engagement. In contrast, topics like "Traditional Culture," "Self-Improvement," and "Popular Culture" are less bursty, showing more stable and consistent engagement patterns. The results are consistent across both consumption and contribution activities. Notably, these most (least) bursty topics

appeared on the platform before (after) Zhihu started open registration. These findings suggest that topics shared by a small community tend to appear or disappear more rapidly than those shared by a wide audience.

6.3 Individual Consumption Preference Dynamics

We now turn to the evolution of individual users' consumption preferences by tracking $\beta_{imk}^{t_i}$ (where $m = s$) across topics for each user i over their lifetime.

6.3.1 Preference clusters

Given the large number of users under study, we start by clustering users into a few segments for data visualization. Each user i 's consumption interests at lifetime t_i are represented as a 20-dimensional embedding $\beta_{isk}^{t_i} / \sum_{k'=1}^K \beta_{isk'}^{t_i}$, where each component k captures the relative intensity on topic k . Our procedure proceeds by first partitioning each user's lifespan into five semiannual periods, $p \in \{1, 2, 3, 4, 5\}$, and then computing the average semiannual embedding for each user. The goal is to derive a sequence of five successive embeddings that capture the evolution of user interests over time. For better visualization, we apply Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018; Dew, 2023) to project user embeddings on a two-dimensional plane. Specifically, we take user embeddings from all five semiannual periods as collective inputs for the UMAP algorithm to ensure that all the maps are oriented consistently and users' positions across time remain comparable. We include 17 anchor tags by incorporating their topic weight vectors (captured in the loading matrix ϕ) as additional input points to the UMAP algorithm.⁷ This approach allows us to infer users' domain interests based on their proximity to these anchor tags.

For ease of interpretation, we cluster all users' embeddings over the five periods into four clusters using the K-means algorithm. This ensures consistent cluster positions across periods, so any observed shifts in users' cluster affiliations signify a change in their interests. Figure 8 shows the resulting clusters, with the dotted lines as the cluster boundaries and the nodes as the cluster centroids over time. We find that the entire space can be conceptualized as a spectrum, with hard-topic tags (such as Finance and Economics, Science and Technology, and Healthcare) predominately in the lower-left fringe, while soft-topic tags (such as Travel, Design, and Sports) mainly in the upper-right corner. Accordingly, we name the lower-left and upper-right clusters as *Hard Enthusiasts (HE)* and *Soft Enthusiasts (SE)*. The two central clusters exhibit

⁷These tags are Internet, Sports, Healthcare, Business, Psychology & Relationships, Education, Travel, Law, Life, Science & Technology, Food, Career, Arts, Design, Finance & Economy, Reading & Writing, Music & Films & Games. Details on their embeddings are provided in Web Appendix C.

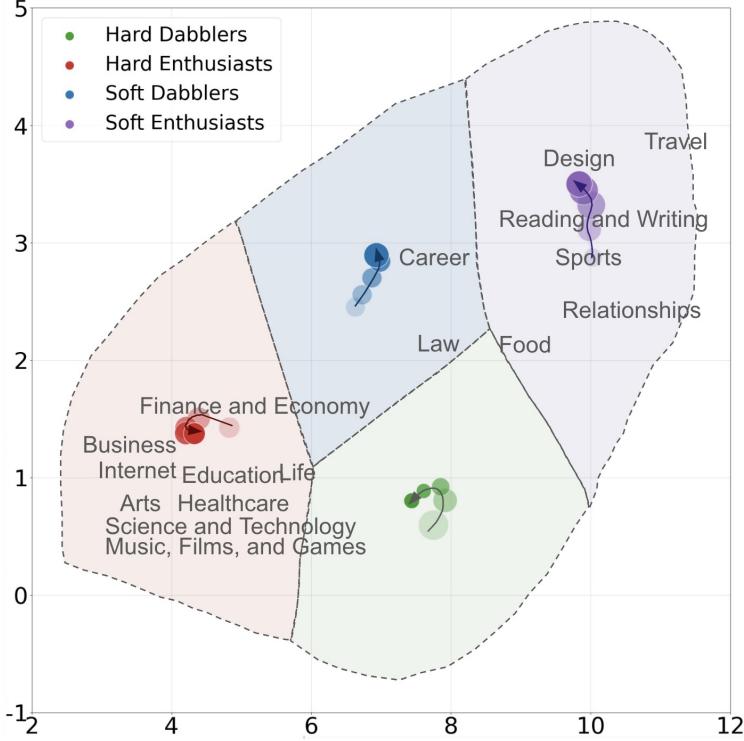


Figure 8: Visualizing the Evolution of Individual Consumption Preferences.

Note: The dotted lines indicate the boundaries of each cluster. The nodes represent the cluster centroids, with their sizes reflecting the number of users in each cluster. Lighter node colors denote earlier periods. The arrows illustrate the movement of cluster centroids over time.

ambiguous and mixed interests. We name them as *Hard Dabblers (HD)* and *Soft Dabblers (SD)*, according to their inclination toward hard and soft topics over periods.

6.3.2 Preference migration

Table 3 reports the proportions of the four user clusters across three semiannual periods. Initially, 62.1% of users were classified as either *Hard Dabblers* or *Soft Dabblers*, but this figure dropped to 42.8% by the fifth period, indicating that most users began with vague interests but developed more defined preferences over time. Meanwhile, the *Hard Enthusiasts* group remained stable, while *Soft Enthusiasts* grew significantly from 16.8% to 34.0%, suggesting a strong growing preference for soft content over time.

We further examine how individual users' cluster affiliations change over time. Insights into user migration patterns can help platforms identify growth opportunities. Figure 9 presents the proportion of users who belonged to each cluster in the first semiannual period and migrated to each cluster in the third and fifth periods. We find that users who were initially *Hard Enthusiasts* exhibited high cluster loyalty, with 41.3%

Table 3: Consumption-Based Cluster Size over Periods.

Clusters	1st Period	3rd Period	5th Period
Hard Enthusiasts	0.210	0.253	0.232
Hard Dabblers	0.438	0.161	0.120
Soft Dabblers	0.183	0.191	0.308
Soft Enthusiasts	0.168	0.395	0.340

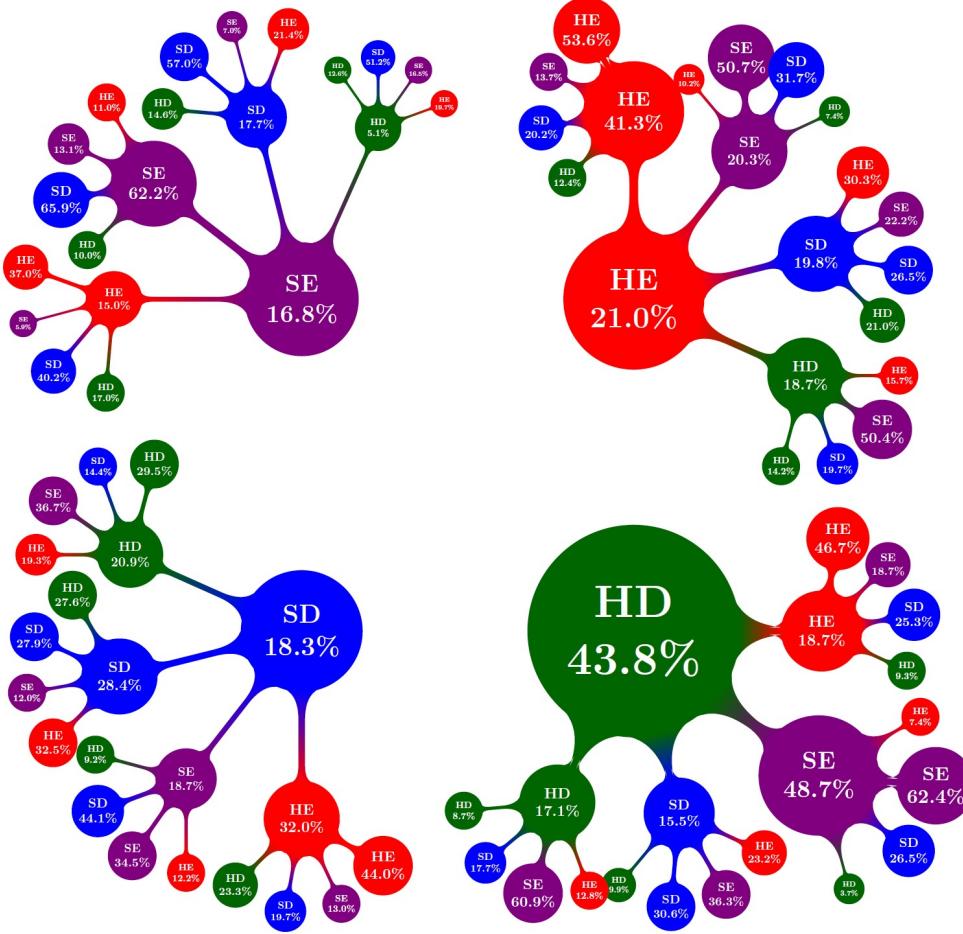


Figure 9: User Migration Patterns across Consumption-Based Clusters.

Note: We code each cluster as SE = Soft Enthusiasts, HE = Hard Enthusiasts, SD = Soft Dabblers, HD = Hard Dabblers. The root of each tree represents users belonging to a cluster in the first semiannual period, while the second and third levels correspond to user cluster assignments in the third and fifth semiannual periods, respectively.

remaining in the same cluster in the third period and $41.3\% \times 53.6\% = 22.1\%$ still in this cluster by the fifth period. This indicates a strong commitment to specialized topics among users with initially clear and focused interests in hard topics. In contrast, users initially classified as *Soft Dabblers* and *Hard Dabblers* ex-

hibited more fluid cluster affiliations. Notably, 48.7% of *Hard Dabblers* transitioned to the *Soft Enthusiasts* cluster by the third period, and $18.7\% + 48.7\% + 15.5\% = 98.2\%$ of them never returned to their original cluster. This pattern suggests that users with initially vague or mixed interests are more prone to explore and eventually settle in domains related to soft topics.

These findings have broad managerial implications for content targeting. First, users' initial interests greatly impact their future content consumption and learning paths, suggesting that platforms should leverage this information to cater to users' evolving content needs. Users with clear initial interests in hard topics can be targeted with specialized and in-depth content, while those with vague initial interests may benefit from a broader range of content to help refine their preferences. In addition, recognizing that a significant portion of users start as dabblers, platforms can initially offer diverse content and focus recommendations as users' interests become clearer. Encouraging dabblers to explore and eventually commit to specific topics can lead to a more engaged and loyal user base, ultimately driving platform growth. For example, users showing initial interest in multiple hard topics can be nudged towards deeper engagement in one specific area. Lastly, platforms should prioritize promoting soft content due to its increasing appeal over time, particularly for users who start with ambiguous interests.

6.4 Individual Contribution Preference Dynamics

6.4.1 Preference clusters and migration

We now examine users' contribution preferences $\{\beta_{iak}^{t_i}\}$ using similar steps. The resulting UMAP projection, as shown in Web Appendix C, shares the same types of user clusters as those defined based on consumption preferences. The overall patterns are similar to individual consumption preferences, so we only highlight a few interesting distinctions. We find that the proportion of *Hard Enthusiasts* dropped significantly from 27% to 19% over the observed periods (see Table W3). This indicates that while the number of users who consume hard topics remained relatively stable, the number of users contributing to these topics decreased substantially. In addition, we find that among users who initially classified as *Hard Enthusiasts* for content contribution, 41.5% remained in the same cluster in the third period, but only $41.5\% \times 36.5\% = 15.1\%$ was still in this cluster by the fifth period (see Figure W8). This percentage is notably lower than that for content consumption in Figure 9 (22.1%). The decline in *Hard Enthusiasts* for content contribution suggests that consistently contributing to hard topics is relatively rare. This is likely

associated with increasing interest in consuming soft topics over time, which diminishes the visibility of hard topics and thus reduces the motivation of experts to contribute. Therefore, it is important for platforms to identify these hard-topic contributors who are at risk early on, and then offer targeted interventions to keep their contributions.

For interested readers, we visualize the estimated four transition matrices of the correlated HMM that capture the evolution of individual preferences. Overall, we find similar migration patterns between hard and soft topics. Details are presented in Web Appendix C.

6.4.2 The diversity of individual preferences

The above results imply that users who remained in the *Hard Enthusiasts* cluster by the fifth period were especially dedicated to hard domains. This is further confirmed by analyzing the diversity of each user's contribution interests. We compute the Herfindahl-Hirschman Index (HHI) (Rhoades, 1993) at each user i 's lifetime t_i as $HHI_i^{t_i} = \sum_k (100 \times (\beta_{iak}^{t_i} / \sum_{k'=1}^K \beta_{iak'}^{t_i}))^2$. A higher value indicates less diverse contribution interests across topics.

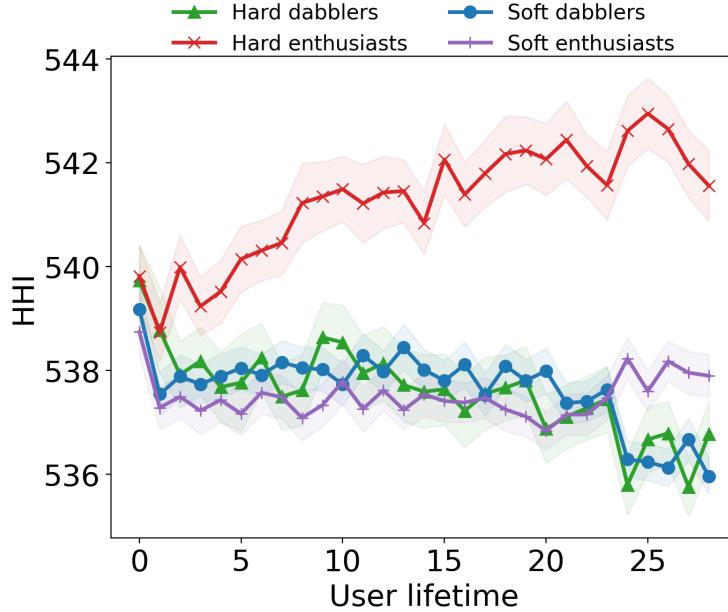


Figure 10: Diversity of User Contribution Preferences over Time by Eventual Cluster.

Note: The lines are the average HHI across users who ended up in the same cluster in the last period, and the shaded regions indicate standard errors.

We created four user groups based on their contribution clusters in the last period. Figure 10 presents the

average HHI within each group over time. We find that the HHI of these users who eventually were *Hard Enthusiasts* increased over time. These users remained dedicated to hard topics and focused their contributions more narrowly on specific areas of expertise as time progressed. Conversely, users who eventually belonged to the *Soft Enthusiasts* cluster maintained stable contribution diversities over time. The remaining two user clusters diversified their contributions over time, expanding to a broader range of topics over time. In the subsequent section, we examine how these different content contribution patterns relate to user follower size, providing insights into how various contribution strategies influence user popularity and reputation.

6.5 Individual Consumption-Contribution Discrepancy

6.5.1 Discrete metrics

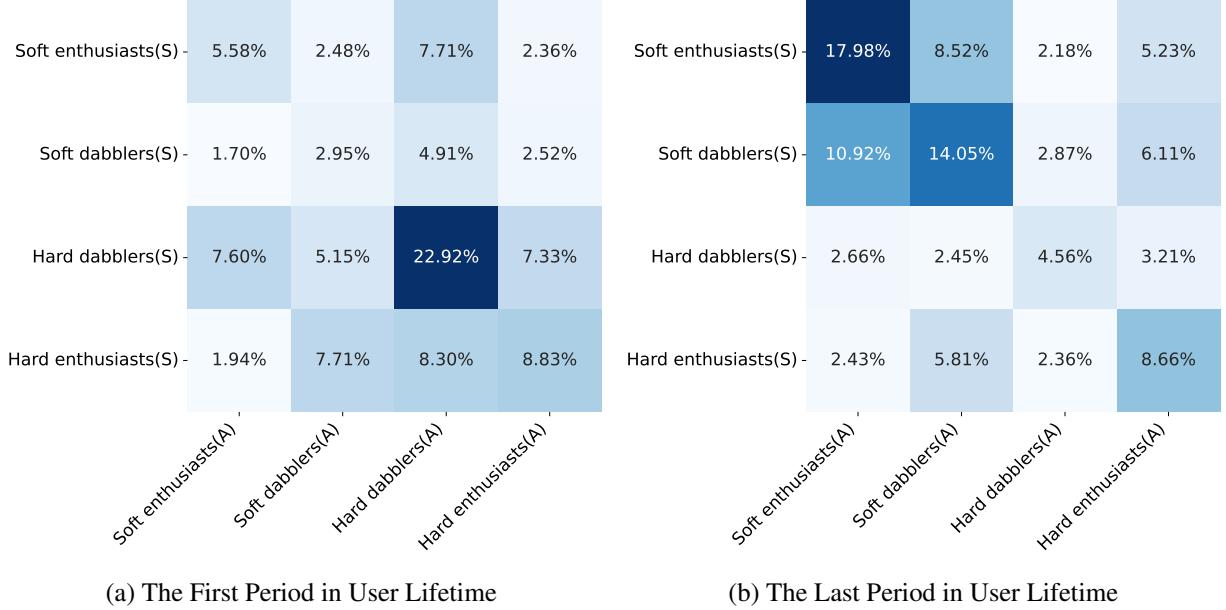


Figure 11: The Proportions of Different Consumption-Contribution Subgroups

Note: (S) denotes subscribing questions, and (A) denotes answering questions.

We now consider individual users' preferences for consumption and contribution simultaneously. Based on the previously defined user clusters, there are $4 \times 4 = 16$ possible consumption (S)-contribution (A) combinations. Figure 11 displays the distribution of these combinations at the first and last semiannual periods of user lifetime, respectively. In each 4-by-4 matrix, the diagonal elements denote the proportions of users who belong to the same type of clusters for consumption and contribution; the upper-diagonal

(lower-diagonal) elements denote the proportions of users whose contribution is relatively more (less) hard-oriented than their consumption.

During the first semiannual period, most users (60%) initially had different consumption and contribution preferences. The most dominant combination, comprising 23% of users, was *Hard Dabblers* for both consumption and contribution, while other diagonal combinations ranged from 3% to 9%. However, by the last semiannual period, there was a high increase in users both contributing to and consuming soft topics. Notably, 47% of *Hard Dabblers* in the first period transitioned to *Soft Enthusiasts* for contribution, and 52% transitioned to *Soft Enthusiasts* for consumption. Additionally, the group size for those contributing to hard topics while consuming soft topics doubled. For example, the new members of the *Soft Enthusiasts(S)* - *Hard Enthusiasts(A)* group primarily came from *Hard Dabblers(S)* - *Hard Dabblers(A)* (32%) and *Hard Dabblers(S)* - *Hard Enthusiasts(A)* (16%) in the first period. In contrast, the group consistently contributing to and consuming hard topics showed no change in size.

6.5.2 Continuous metrics

We also construct a continuous metric to measure consumption-contribution discrepancy using the L_2 norm, i.e., $\|\beta_{ia}^{t_i} - \beta_{is}^{t_i}\|_2$. Figure 12 displays the average discrepancy across all users over time. We find an inverted U-shaped pattern with a peak of 2.90 in the 13th month of users' lifetimes. This suggests that user consumption and contribution preferences initially diverged but gradually synchronized as users spent more time on the platform. However, different patterns emerge when examining these metrics across the 16 consumption-contribution groups based on a user's eventual membership in the last period. Figure 13 illustrates these patterns, with the left panel showing groups with increasing discrepancy and the right panel showing those with decreasing discrepancy. Overall, users who either contributed or consumed the hard topics exhibited larger and growing discrepancies, but the remaining groups exhibited smaller discrepancies that declined over time.

These findings have important implications. First, users with an affinity for hard topics tend to develop divergent consumption and contribution preferences over time, potentially due to a growing tendency to contribute hard content while consuming softer content. Conversely, users interested in soft topics tend to align their consumption and contribution preferences more closely, likely because soft content is less cognitively demanding, allowing for consistent engagement across both activities. These results underscore the need for activity-dependent content promotion and the importance of recognizing potential discrepancies

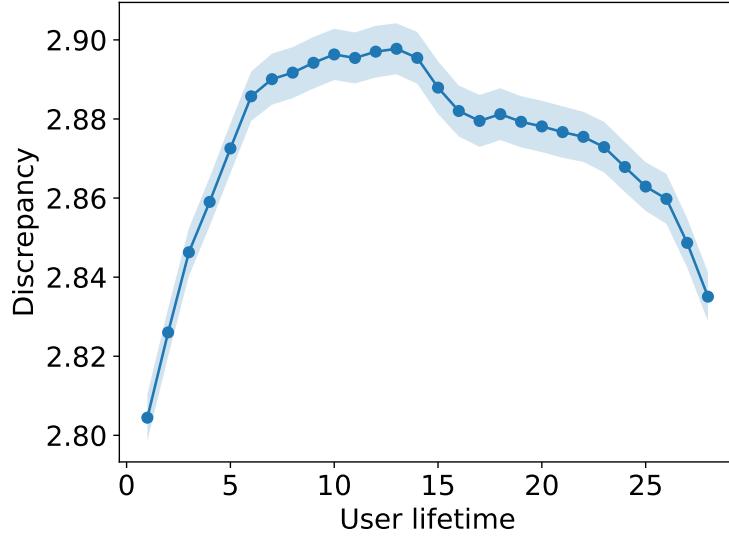


Figure 12: Consumption-Contribution Discrepancy over Time.

Note: The line is the population average, and the shaded region indicates the standard error.

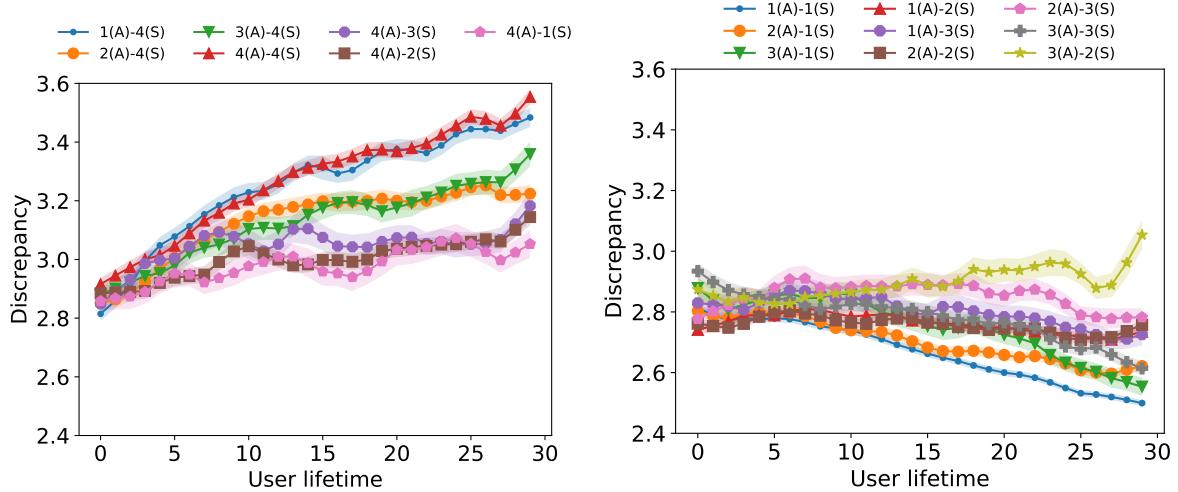


Figure 13: Consumption-Contribution Discrepancy.

Note: The lines represent the average discrepancy level among users who ended up in the same cluster in the final period, with shaded regions showing the standard errors. The degree of hard topic preference is coded as follows: 1 = Soft Enthusiasts, 2 = Soft Dabblers, 3 = Hard Dabblers, and 4 = Hard Enthusiasts. (S) and (A) denote subscription and answering activities, respectively.

between users' consumption and contribution behaviors. Platforms should consider offering a broader range of content to users interested in hard topics, though such strategy may be less effective for those interested in general and soft topics.

6.6 The Decomposition of Individual Activities

This subsection investigates to what extent user activities are associated with individual preferences (i.e., $\beta_{imk}^{t_i}$) relative to the aggregated platform trends (i.e., $\theta_{mk}^{t_i^0 + t_i}$). We approximate this relationship using a ratio: $R_{imk}^{t_i} = \beta_{imk}^{t_i} / \theta_{mk}^{t_i^0 + t_i}$ for $m \in \{a, s\}$. For ease of illustration, we compute the average $R_{imk}^{t_i}$ across time for each user-topic-activity combination, denoted by \bar{R}_{imk} . We find that \bar{R}_{imk} exceeds one in over 90% of observations, indicating that user activities are more strongly linked to individual preferences than to aggregate trends. Moreover, the average \bar{R}_{imk} is significantly higher for question subscription than for answer contribution ($p < 0.01$), suggesting that contribution aligns more closely with platform trends than consumption.

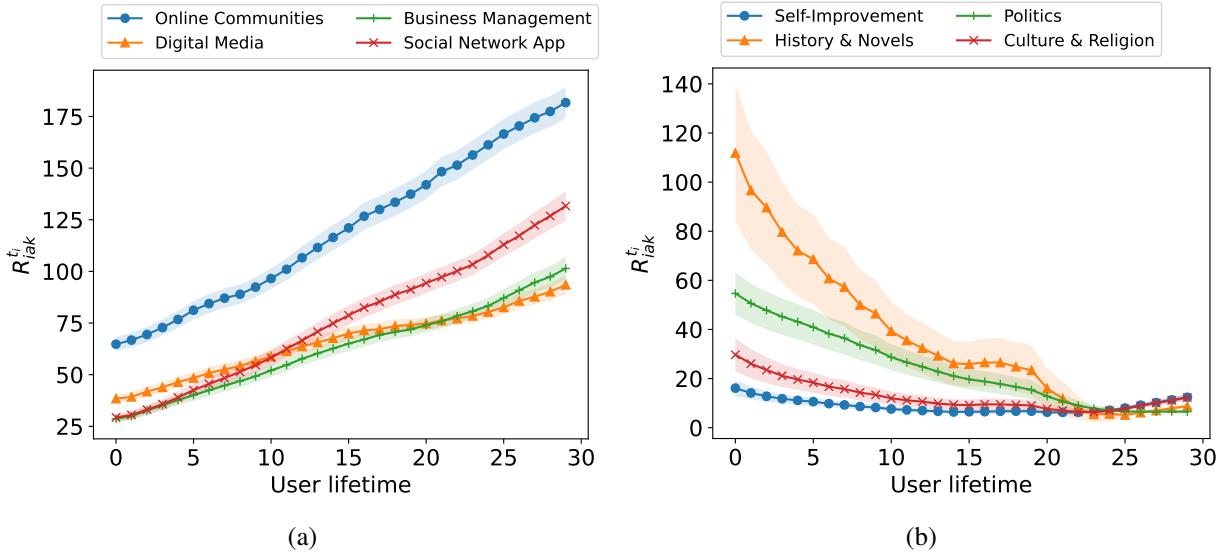


Figure 14: The Evolution of Individual-Platform Activity Ratio for Selected Topics.

Note: The ratio $R_{iak}^{t_i} := \beta_{iak}^{t_i} / \theta_{ak}^{t_i^0 + t_i}$. Thus, $R_{imk}^{t_i} > 1$ indicates that user intrinsic preference has a relatively stronger influence on user activity than the general platform trends.

We further explore the heterogeneity across topics. Figure 14 displays the average contribution-based ratio across users over their lifetime for eight sample topics. Results for all topics are provided in Web Appendix C. There are notable increasing trends for several topics, such as “Online Communities” and “Social Network App,” indicating sustained and inherent user interests in IT-related areas. This aligns with the fact that many active users in our study window are IT professionals. In contrast, general-interest topics, such as “Politics” and “Culture & Religion,” show declining trends. This indicates that user interests in general-interest topics are more likely to be associated with platform trends than personal preferences.

6.7 Preferences and Follower Sizes

These estimated individual preferences allow us to explore how user preferences are associated with other individual outcomes of interest. As an example for demonstration, we consider each user's follower size at the end of her lifetime, denoted by $Follower_i$. Follower size is the most important indicator of user success on digital content platforms and is frequently utilized by platforms to determine various decisions, e.g., identifying content experts and promoting users.

Table 4: Regression of Follower Size on User Eventual Cluster Affiliations

Dep. Var.	$Ln(Follower + 1)$		
	(1)	(2)	(3)
<i>SoftDabblers</i> ^A	0.343*** (0.051)		0.295*** (0.052)
<i>HardDabblers</i> ^A		1.256*** (0.069)	1.021*** (0.073)
<i>HardEnthusiasts</i> ^A		2.114*** (0.060)	1.889*** (0.064)
<i>SoftDabblers</i> ^S		0.379*** (0.054)	0.117** (0.054)
<i>HardDabblers</i> ^S		1.154*** (0.073)	0.632*** (0.075)
<i>HardEnthusiasts</i> ^S		1.230*** (0.058)	0.616*** (0.060)
<i>Intercept</i>	3.172*** (0.036)	3.317*** (0.037)	3.007*** (0.042)
Adj. R-sq.	0.087	0.036	0.096

Note: The table reports regressions of log-transformed follower size on user contribution- and consumption-based cluster affiliation at the last semiannual period. The follower size is the cumulative follower count of each user by the end of their 30th month on the platform. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

We first regress the log-transformed $Follower_i$ on user i 's contribution- and consumption-based cluster affiliations observed at the last semiannual period. Table 4 presents the estimation results, while setting the *Soft Enthusiasts* cluster as the baseline group. All model coefficients are significantly positive($p < 0.01$), with *Hard Enthusiasts* showing the highest follower size, followed by *Hard Dabblers*, *Soft Dabblers*, and *Soft Enthusiasts*. This suggests that users with hardcore expertise/interests tend to attract more followers than users with soft-oriented preferences. Intuitively, we find that contribution preferences are more strongly related to user follower size than consumption preferences based on the adjusted R-square values and the magnitudes of the coefficients. We also replicated this analysis but using user cluster affiliations at the first and third semiannual periods instead. Estimation results are reported in Appendix C. We find that

preferences revealed early in user lifetimes have weak effects on follower size. This finding confirms that users need time to explore and form their preferences/expertise on the platform.

Table 5: Relationship Between R_{imk} and Follower Size .

Topics	Estimated Coefficients			
	\bar{R}_{iak}	$\bar{\beta}_{iak}$	\bar{R}_{isk}	$\bar{\beta}_{isk}$
Mobile Device	-0.007*	0.782**	-0.002	0.350
Traditional Culture	-0.001	0.601***	0.001	-0.164
Sciences & Engineering	0.011	0.362**	0.002	0.163
Online Communities	-0.002	0.182	0.000	-0.072
Digital Media	-0.002	0.326	0.001*	-0.115
Lifestyle	0.035	0.248	-0.001	0.268*
SEO-Marketing	-0.007*	0.783**	0.001	-0.210
E-Commerce	-0.001	0.130	0.000**	0.314*
Entertainment & Trends	-0.004***	0.366***	-0.001	0.152
Zhihu Platform	0.001	0.012	-0.001	0.112
Others	-0.007	0.355	-0.001*	-0.124
Social Network App	-0.001	0.277	-0.001	-0.201
Tech-Finance	-0.003**	0.484**	0.000	0.270
Self-Improvement	0.037**	0.246	-0.001	0.239
Popular Culture	0.045	0.102	0.004	-0.067
Relationships	0.000	0.955***	0.000***	-0.420***
Politics	0.004	0.413**	0.000	-0.112
Culture & Religion	0.002	0.585***	0.000	0.097
History & Novels	0.003*	0.572***	0.000	0.162
Business Management	-0.018***	0.884***	0.000	0.316**

Note: The table reports the point estimates of a regression model of log-transformed $Follower_i$ on 80 preference-related variables, including the average individual preferences $\bar{\beta}_{iak}$ and the ratio metric \bar{R}_{imk} defined previously for 20 topics and two action types (i.e., consumption (s) and contribution (a)). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Next, we quantify the relationship between follower size and the continuous measures of individual preferences and the ratio metric \bar{R}_{imk} defined previously. We specify the log-transformed $Follower_i$ as

$$\ln(Follower_i + 1) = \alpha + \sum_{m \in \{s,a\}} \sum_{k=1}^K (\gamma_{mk} \bar{R}_{imk} + \eta_{mk} \bar{\beta}_{imk}) + \varepsilon_i, \quad (9)$$

where $\bar{\beta}_{imk}$ denotes the average $\beta_{imk}^{t_i}$ across times for each user-topic-action combination. Thus, this model contains 80 independent variables that capture user consumption and contribution interest across topics. The

model fits the data reasonably well, with an R-square of 0.344. The point estimates are shown in Table 5, in which the first (last) two columns are variables related to content contribution (consumption). Notably, the coefficients of $\bar{\beta}_{iak}$ are uniformly positive, with 11 of them being statistically significant ($p < 0.05$), whereas the majority of the coefficients of \bar{R}_{iak} are negative. These findings suggest that while active content contribution can increase user popularity, diverging significantly from the prevailing trends on the platform may not yield the desired outcomes. Creators with specialized and niche expertise should mindfully design the content and style of their contributions according to the general trends on the platform. As expected, the coefficients of consumption-related variables are less significant and have smaller magnitudes than those of contribution-related variables. This further confirms that contribution preferences are more important to user success than consumption preferences.

7 Discussion and Conclusion

Given the importance of enhancing user engagement on digital content platforms, this paper proposes MDNPS to understand user preferences for unstructured content. The MDNPS can unravel multifaceted consumer preferences, from the lens of different user activities, time periods, and granularity levels. It yields high model flexibility, interpretability, and scalability by combining Bayesian principle with neural networks in a deep dynamic Poisson factor model. MDNPS reveals interpretable topics and hierarchical user preferences over these topics at various times and activities. For efficient model inference, we propose a variational Bayes algorithm to jointly learn the neural network parameters and infer the posterior distribution of latent topics and the preferences over the topics. This model is scalable to large datasets involving high-dimensional and diverse user activities. We conduct a comprehensive empirical application to Zhihu users. MDNPS significantly improves in-sample and out-sample data fitting compared to the existing models.

Substantively, MDNPS captures platform- and individual-level dynamics, offering rich managerial implications for digital content platforms and content creators. At the platform level, the model effectively summarizes users' sparse interactions with tags into semantically coherent topics that reflect the platform's topical spectrum. The model captures the dynamics of collective user interests across these topics, which closely synchronize major external events or societal trends. The model identifies the topic-level demand-supply gaps, offering pathways for platforms to encourage expert contributions where needed most, enhancing content relevance and user satisfaction. Moreover, it extracts detailed properties of topic dynamics, such

as persistence, emergence, and burstiness, to facilitate content management. At the individual level, the model characterizes clear trajectories that users typically follow and segments users into four major clusters. We find that users' initial interests and expertise can inform the patterns of user preferences evolution. Furthermore, we analyzed the relationships between user preferences and follower size. We find that contributing to hard topics tends to be associated with a larger follower size than contributing to (consuming) other topics (any topics). In addition, while user activities are predominantly driven by intrinsic preferences, aligning with platform trends, especially in content generation, can significantly impact user success.

Our findings highlight the importance of adaptive content strategies and user-centric approaches in managing digital platforms. By leveraging user data to understand the nuanced differences across user segments and the dynamic nature of user interests, platforms can better tailor their content offerings, fostering sustained engagement and community growth. Future research can further inform best practices and theoretical frameworks. In addition, our study confirms the importance of understanding the effectiveness of different personalization algorithms in catering to diverse user interests. Research can focus on the balance between promoting popular content and surfacing niche topics to meet individual preferences. More broadly, the unraveled latent information from digital platforms enables the collection of public opinions on policies and societal events (West et al., 2021). We view these as fertile areas for future exploration.

Our study has several limitations and suggests important avenues for future research. First, we focus on user interactions with tags to measure preferences. However, in practice, contextual information such as the text description of the tags and characteristics of users (e.g., social networks) might be available. Future studies can incorporate such information as additional input to the LSTM to encode user heterogeneity and tag similarity. Second, beyond the interaction with tags, our model can be seamlessly applied to users' interaction with text documents when the texts are represented as bag-of-words. Third, our current model focuses on associations, while the uncovered latent factors may offer insights into building structural models (Ruiz et al., 2020). For instance, the inferred user preferences can be utilized in modeling the utility function of a consumer demand model, which describes how users allocate their time across various content. Finally, our current model is applied to a single digital content platform. Another dimension where users may exhibit preference heterogeneity is across different platforms. Future work could extend our Bayesian deep learning framework to other digital or offline settings that involve complex, dynamic, and multifaceted user behaviors.

Web Appendices for

“Unraveling Multifaceted User Preferences on Content Platforms: A Bayesian Deep Learning Approach”

A Data Statistics

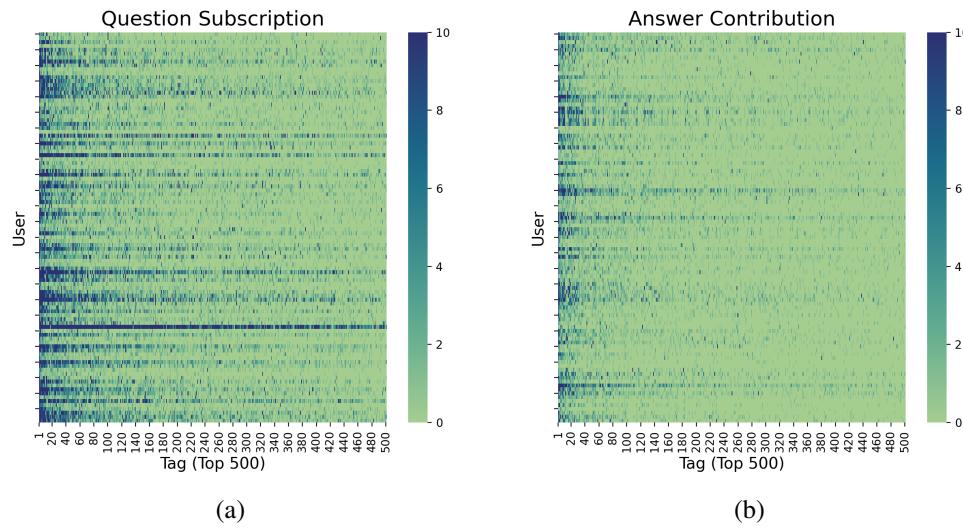


Figure W1: Sample Users’ Activity Frequencies across Tags and Action Types

Note: This table shows the frequency of user interactions (answers or subscriptions) with tags on Zhihu over a 30-month period. It includes data from 100 randomly selected users (y-axis) and the top 500 popular tags (x-axis) during the study period.

B Derivation of the Inference Algorithm

B.1 Variational Distributions

We introduce a set of variational distributions $q_\psi(\beta, \theta, \phi)$ with variational parameters ψ to approximate the unknown posterior distribution. The full variational distribution is designed as

$$q_\psi(\beta, \theta, \phi) = \prod_{t=1}^T \left[q(\theta^t | \theta^{t-1}, \tilde{y}^{1:t}) \prod_{k=1}^K \left[\prod_{i=1}^N q(\beta_{ik}^t) \prod_{j=1}^J q(\phi_{jk}) \right] \right]. \quad (\text{B1})$$

$q(\theta^t | \theta^{t-1}, \tilde{y}^{1:t})$ is constructed with the LSTM neural network to encode the long-term dependences, and \tilde{y} is the bag-of-words representation of the tags at the time t . Instead of using a separate set of variational parameters for each data point, the LSTM maps data to the variational parameters of $\theta_{mk}^{1:T}$. This construction allows sharing the LSTM parameters across time t to improve the computational efficiency (Kingma and Welling 2013). We compute the bag-of-words representation $\tilde{y}^t \in \mathbb{Z}^J$, where $\tilde{y}_j^t = \sum_i y_{ij}^t / n^t$ is the average individual interaction numbers with the tag j , and n^t is the number of active users at time t .

Similar to Equation (??), the LSTM $g(\cdot)$ takes $(\tilde{y}^1, \dots, \tilde{y}^T)$ as the input and outputs the hidden states $(h_q^1, h_q^2, \dots, h_q^T) = g(\tilde{y}^1, \dots, \tilde{y}^T)$. The samples of θ^t are then generated by

$$\theta_m^t \sim q(\theta_m^t | \theta_m^{t-1}, \tilde{y}^{1:t}) = \mathcal{LN}(\mu_q(\theta_m^{t-1}, h_q^t(\tilde{y}^{1:t})), \sigma_q^2(\theta_m^{t-1}, h_q^t(\tilde{y}^{1:t}))I), \quad (\text{B2})$$

where $\mu_q(\cdot), \sigma_q^2(\cdot)$ are two feedforward neural networks. The log-normal distribution is *reparameterizable*, i.e., a sample θ_m^t can be generated by a deterministic transformation of a sample from the standard normal distribution $\mathcal{N}(0, 1)$. This property allows using the reparameterization trick (Kingma and Welling 2013) to efficiently obtain an unbiased gradient estimation of the ELBO without computing the intractable expectation in Equation (B5).

We use mean-field distributions to approximate the posteriors of the individual latent scores β and the factor loading matrix ϕ ,

$$q(\tilde{\beta}_{imk}^t) = \mathcal{N}((\mu_\beta)_{imk}^t, (\sigma_\beta^2)_{imk}^t) \text{ and } \beta_{imk}^t = \exp(\tilde{\beta}_{imk}^t); \quad (\text{B3})$$

$$q(\tilde{\phi}_{jk}) = \mathcal{N}((\mu_\phi)_{jk}, (\sigma_\phi^2)_{jk}) \text{ and } \phi_k = \text{Softmax}(\tilde{\phi}_k). \quad (\text{B4})$$

In a nutshell, the variational parameter ψ consists of the mean-field parameters $(\mu_\beta, \sigma_\beta^2, \mu_\phi, \sigma_\phi^2)$, the parameters of the LSTM $g(\cdot)$, and the parameters of the neural networks $\mu_q(\cdot), \sigma_q^2(\cdot)$.

B.2 Evidence Lower Bound

With the joint distribution $p(y, \beta, \theta, \phi; \eta)$ and variational distribution $q_\psi(\beta, \theta, \phi)$, the objective of variational EM is to maximize the ELBO with respect to the variational parameters ψ and model parameters η

(Blei et al. 2017),

$$\mathcal{L}(\psi, \eta) = \mathbb{E}_{q_\psi(\beta, \theta, \phi)} \left[\log p(y, \beta, \theta, \phi; \eta) - \log q_\psi(\beta, \theta, \phi) \right]. \quad (\text{B5})$$

We take the reparametrization trick (Kingma and Welling 2013) and stochastic variational inference (SVI) technique (Hoffman et al. 2013) to achieve feasible and efficient computation of the ELBO. In particular, the reparametrization trick allows estimating the intractable expectation in Equation (B5) using random samples from $q_\psi(\beta, \theta, \phi)$ ¹. Our variational distributions $q_\psi(\beta, \theta, \phi)$ are designed to satisfy this reparametrization property.

Since the entire data set is large and computationally expensive, we randomly sample a minibatch of B individuals from $\{1, 2, \dots, N\}$ to compute the unbiased minibatch ELBO $\hat{\mathcal{L}}(\psi, \eta)$ with $\mathbb{E}[\hat{\mathcal{L}}(\psi, \eta)] = \mathcal{L}(\psi, \eta)$, and compute the stochastic gradient using $\hat{\mathcal{L}}$ at each round of parameter update. The detailed functional form of the mini-batch ELBO is presented below.

$$\begin{aligned} \hat{\mathcal{L}}(\psi, \eta) &= \mathbb{E}_q \left[\frac{N}{B} \sum_{i \in \text{batch}} \log p(y_i | \beta, \theta, \phi) \right] - \sum_{m=1}^M \left(\sum_{t=1}^T \text{KL}(q(\log(\theta_m^t) | \theta_m^{1:t-1}) || p(\log(\theta_m^t) | \theta_m^{1:t-1}; W_\theta)) \right. \\ &\quad \left. - \sum_{i=1}^N \sum_{t=1}^T \text{KL}(q(\log(\beta_{im}^t)) || p(\log(\beta_{im}^t) | \beta_{i,1:m}^{t-1}; \sigma_0, \Pi)) \right) - \sum_{k=1}^K \text{KL}(q(\tilde{\phi}_k) || p(\tilde{\phi}_k)). \end{aligned} \quad (\text{B6})$$

The first expectation term can be estimated with Monte Carlo samples from $q_\psi(\beta, \theta, \phi)$ using a reparametrization trick (Kingma and Welling 2013). All the other terms are analytic based on the KL divergence between two k-dimensional multivariate Gaussian distributions,

$$\text{KL}(\mathcal{N}(\mu_1, \Sigma_1) || \mathcal{N}(\mu_2, \Sigma_2)) = \frac{1}{2} \left(\text{tr}(\Sigma_2^{-1} \Sigma_1) + (\mu_2 - \mu_1)^T \Sigma_2^{-1} (\mu_2 - \mu_1) - k - \ln \left(\frac{\det(\Sigma_1)}{\det(\Sigma_2)} \right) \right).$$

B.3 Orthogonal Temporal Regularization

Let u_j^t denote the estimated platform-level user interaction frequency with tag j at time t , which is computed as $u_j^t = \sum_k \theta_k^t \phi_{jk}$. We find the cosine similarity between the platform-level estimated tag counts u^t and $u^{t'} \in \mathbb{R}^J$ at two different time points, computed as $u_j^t = \sum_k \theta_k^t \phi_{jk}$, is close to 1 when the topic collapses.

The design of Orthogonal Temporal Regularization (OTR) is motivated by the observed data. To explore

¹The random samples are generated as deterministic mapping of samples from a base distribution such as the standard normal distribution $N(0, I)$.

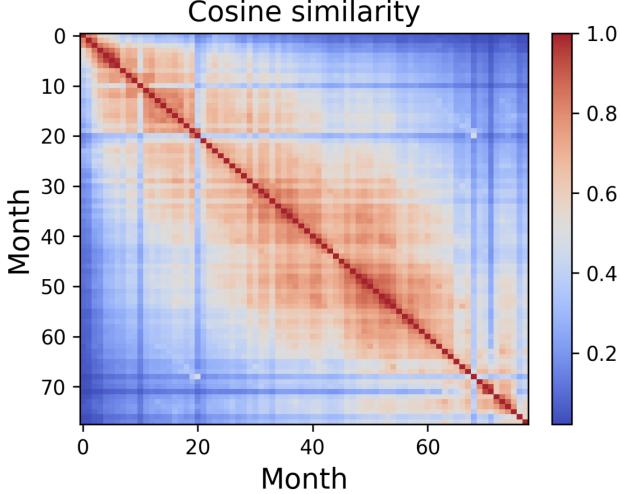


Figure W2: The Average Cosine Similarity between Monthly Counts

Note: We compute the total activity count across all of the individuals for each tag at each month, giving us a vector of activity counts across all tags at each month. Then we compute the cosine similarity of these vectors between all possible pairs of months in user lifetime, t and t' , observed in our data.

what is a reasonable value of cosine similarity between u^t and $u^{t'}$, we approximate u_j^t by the actual interaction frequency \tilde{y}_j^t , which is computed as $\tilde{y}_j^t = \sum_i y_{ij}^t / n^t$ with n^t as the number of active users at time t . We then compute the cosine similarity of \tilde{y}^t and $\tilde{y}^{t'}$ as shown in Figure W2, and check its magnitude. We notice that the cosine similarity of the tag counts is high when t and t' are close but low when t and t' are far apart, with an average cosine similarity of 0.38. In contrast, when the topics collapse, the cosine similarity of u^t and $u^{t'}$ is close to 1. This empirical observation suggests that we should regularize the pairwise cosine similarity between u^t and $u^{t'}$ to mitigate the topic collapse problem.

Specifically, we compute the cosine similarity as $\frac{u^t \cdot u^{t'}}{\|u^t\| \cdot \|u^{t'}\|}$, $u^t = \sum_k \theta_k^t \phi_k$ and set the OTR as $\mathcal{R} = \sum_{t < t'} \frac{u^t \cdot u^{t'}}{\|u^t\| \cdot \|u^{t'}\|}$. The objective in Equation (B6) is then modified to obtain our final objective

$$\max_{\psi, \eta} \hat{\mathcal{L}}(\psi, \eta) - \alpha \mathcal{R}, \quad \mathcal{R} = \sum_{t < t'} \frac{u^t \cdot u^{t'}}{\|u^t\| \cdot \|u^{t'}\|}. \quad (\text{B7})$$

Above, $\hat{\mathcal{L}}(\psi, \eta)$ is the mini-batch ELBO and $\alpha \geq 0$ decides the regularization strength. The OTR \mathcal{R} is maximized to $T(T-1)/2$ when the topics collapse; hence, controlling its magnitude would mitigate the topic collapse problem. Unlike existing orthogonal regularization that encourages orthogonality across factors k , we find that it is crucial to encourage orthogonality across time t for our temporal dynamic data.

B.4 Variational Expectation-Maximization Algorithm

The algorithm proceeds by optimizing the regularized ELBO in Equation (B7) with respect to the collection of model parameters η discussed in the “Model Specification” subsection and the variational parameters ψ discussed in the “Evidence Lower Bound” subsection. The variational and model parameters can be iteratively updated by the variational EM algorithm with the stochastic gradient descent (SGD) (Blei et al. 2017). Each iteration consists of an expectation step (E-step) to update ψ and a maximization step (M-step) to update η as

$$\begin{aligned} \text{Variational E-step: } \psi^{\text{new}} &= \psi + \tau_\psi \nabla_\psi (\hat{\mathcal{L}}(\psi, \eta) - \alpha \mathcal{R}); \\ \text{M-step: } \eta^{\text{new}} &= \eta + \tau_\eta \nabla_\eta (\hat{\mathcal{L}}(\psi, \eta) - \alpha \mathcal{R}). \end{aligned} \quad (\text{B8})$$

The algorithm repeats the Variational E-step and M-step until convergence. In our implementation, the stepsizes τ_ψ, τ_η are scheduled by the Adam algorithm (Kingma and Ba 2014), and the SGD updates are implemented by PyTorch (Paszke et al. 2019). The E-step improves the inference by approximating the posterior distribution with the variational distribution, and the M-step improves the data fitting by maximizing the lower bound of the marginal likelihood.

B.5 Model Validation Metrics

The topic coherence (Mimno et al. 2011) is defined as

$$\text{TC} = \frac{1}{K} \sum_{k=1}^K \frac{2}{C(C-1)} \sum_{1 \leq i < j \leq C} f(w_i^k, w_j^k) \quad (\text{B9})$$

where, w_1^k, \dots, w_C^k are the top C tags in topic k with the highest probability; $f(w_i^k, w_j^k)$ is the normalized pointwise mutual information between the word pair (w_i^k, w_j^k) , which is defined as

$$f(w_i^k, w_j^k) = \left[\log \frac{p(w_i^k, w_j^k)}{p(w_i^k)p(w_j^k)} \right] / \left[\log(1/p(w_i^k, w_j^k)) \right]. \quad (\text{B10})$$

Take question answering for example. $p(w_i^k, w_j^k)$ is estimated by the proportion of users whose answers are associated with both tags w_i^k and w_j^k , and $p(w_i^k)$ is estimated by the proportion of users whose answers are associated with tag w_i^k .

The perplexity is defined as

$$\text{Perp}(D_{\text{hold-out}}) = \exp \left(-\frac{\sum_{y_{ij} \in D_{\text{hold-out}}} \log p(y_{ij})}{\sum_{y_{ij} \in D_{\text{hold-out}}} N_i} \right), \quad (\text{B11})$$

where y_{ij} are the action counts over the tags j of the user i at the hold-out time and $N_i = \sum_j y_{ij}$. It measures how well a probability model predicts a hold-out sample.

C Additional Substantive Outputs and Analysis

C.1 Interpretable Topics

Table W1: Top Words in Latent Topics.

Topic	Topwords
<i>Mobile Device</i>	iPad Apps, E-Book, iOS 7, CSS, Magazine, Internet Community Operation, Cloud Computing, Search, iOS App Recommendations, Windows 8
<i>Traditional Culture</i>	Writing Skills, Historical Knowledge, Ancient Poems, History, Evolution, Three Kingdoms, A Song of Ice and Fire (Novel), Smartisan OS, Mobile Photography, Watch
<i>Engineering and Sciences</i>	Agriculture, Civil Engineering, Manufacturing, Writing Skills, Computer Major, China Stock Market, Cancer, Securities, C/C++, Firearms
<i>Online Communities</i>	Steve Jobs, Magazine, Quora, Use Zhihu, Geek, PHP, iPhone App, Team Management, Online Forum, Internet Community
<i>Digital Media</i>	E-reading, Evolution, Publishing, MacBook Air, Code, Personal Computer, Visual Design, CSS, Video Website, Memory
<i>Lifestyle</i>	iOS 7, iOS Games, Efficiency, Entrepreneur, Health Knowledge, Daily Life, MacBook Air, Coffee, American Culture, Team Management
<i>SEO-Marketing</i>	Publishing, Search, Search Engine Optimization (SEO), Magazine, Web Games, E-Book, China Mobile, Team Management, Weibo Marketing, Website Operation
<i>E-Commerce</i>	Internet Product Design, Zhihu Product Improvements, Group Buy, Tencent Weibo, Social Products, B2C E-commerce, Website Operation, User Research, Search Engine Optimization (SEO), Group Buying Website
<i>Entertainment and Trends</i>	Donald J. Trump, Bilibili, Game of Thrones (American TV series), Hearthstone, Hairstyle, Overwatch, A Dream of Red Mansions (Novel), LGBTQIA, PlayStation 4, Love Relationship
<i>Zhihu Platform</i>	Zhihu Product Improvements, iPhone App, Tencent Weibo, Search Engine Optimization (SEO), Social Products, User Research, B2C, Zhihu Guide, Zhihu Design, Website Operation

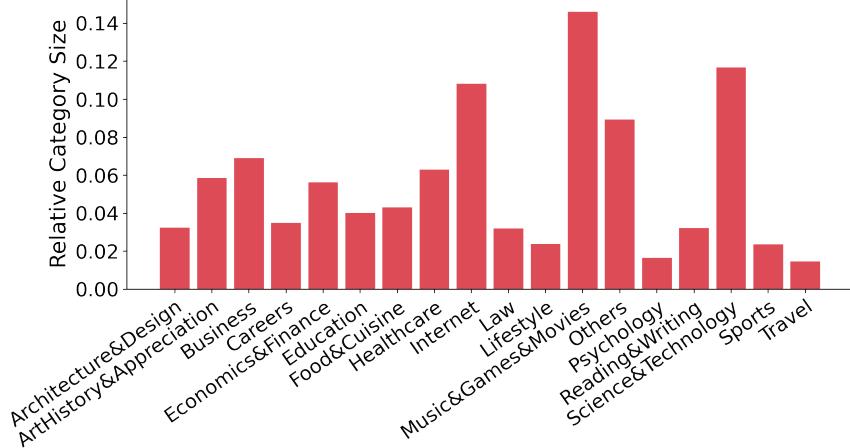
<i>Others</i>	Name, People, FC Barcelona, Disease, Personal Computer, Time Travel, Intellectual Property, Stephen Chow (Character), Meaning of Life, International Students
<i>Social Network App</i>	Sina, Weibo Marketing, Zhihu Guide, Renren, Mobile development, Search Engine Optimization (SEO), Internet Community Operation, Zhihu Design, Visual Design, Social Products
<i>Tech-Finance</i>	Bitcoin, iOS 7, E-book, US Stocks, Smartisan OS, Initial Public Offering (IPO), iPad Air, Wi-Fi, Visual Design, Insurance
<i>Self-Improvement</i>	Investment Bank, Speaking Skills, Civil Engineering, Finance, Hardware, Policy, Game Recommendations, Social Etiquette, Efficiency, Personal Development
<i>Popular Culture</i>	Guo Jingming (Character), Smartisan OS, Chinese Studies, Commodity, Popular Science in Physics, iPhone App, Finance, House of Cards, Chinese Literature, Hardware
<i>Relationships</i>	Appearance, Intimacy, Childhood, Self-Management, Find a Partner, Height, Goddess, Part Time, Life Experience, Social Etiquette
<i>Politics</i>	Donald J. Trump, 2016 Olympic Games in Rio de Janeiro, 2016 US Election, Hillary Clinton, Wang Baoqiang, US Election, International Situation, American Politics, Religious Power, Taiwan Secession Issue
<i>Culture & Religion</i>	Social Etiquette, Goddess, Developmental Psychology, Ancient Poems, Sexual Attraction, Buddhism, Calligraphy, Buddha, Cognition, Tea
<i>History & Novels</i>	Historical Knowledge, Soviet Union, History, Post-90s, A Song of Ice and Fire (Novel), A Dream of Red Mansions (Novel), Writing Skills, Three Kingdoms, ISIS, Rural Area
<i>Business Management</i>	Team Management, Website Operation, E-Commerce Operations, Social Products, Blog, Kai-Fu Lee (Character), Magazine, Smartisan OS, Personal Computer, E-Reading

C.2 Anchor Tags

Following the approach in Liu and Cong (2023), we utilize Zhihu’s topic tree and Graph Convolutional Network to classify tags in our sample into the 17 anchor tags (hereafter “categories”). Figure W3 presents the distribution of tags across these anchor categories.

For visualization, we represent each category as the average embeddings of tags within it. The tag embeddings are derived from the loading matrix ϕ , which contains vectors of weights across the 20 latent topics for each tag. Thus, these category embeddings are in the same space as user embeddings, allowing us to infer each user’s domain interests using her nearby categories.

Figure W3: Proportions of Tags per Category



C.3 Transition Matrices

Our model estimates four transition matrices, Π^{ss} , Π^{as} , Π^{aa} and Π^{sa} , which are presented in Figure W4. The transition matrix Π^{ss} (Π^{as}) encodes how the consumption (contribution) interests at time $t - 1$ transit to the consumption interests at time t . Each matrix has dimension $K \times K$. The k -th row quantifies the transition probability from all K topics at time $t - 1$ to the k -th topics at time t , with the row sum $\sum_{k'=1}^K \Pi_{kk'}^{ms}$ represents the degree of interests transits into the topic k . Similarly, the k -th column quantifies the transition probability from the k -th topic at time $t - 1$ to all K topics at time t , with the column sum $\sum_{k'=1}^K \Pi_{k'k}^{ms}$ represents the degree of interests transits out of the topic k . Accordingly, we define the relative ratio of the k -th row sum to the k -th column sum $R_k^{ms} = (\sum_{k'=1}^K \Pi_{kk'}^{ms} - \sum_{k'=1}^K \Pi_{k'k}^{ms}) / \sum_{k'=1}^K \Pi_{k'k}^{ms}$ as the growth rate of the topic k on the platform over the 78 months. A positive value indicates that the interest flowing out of a topic exceeds the interest it attracts, and a negative value indicates the other way around.

In what follows, we highlight a few examples based on Π^{ss} and Π^{as} to demonstrate how the platform can interpret these matrices and derive managerial insights. Figure W5 visualizes two rows of Π^{ss} about the topics “Zhihu Platform” and “Entertainment and Trends.” We find that the consumption interest in “Zhihu Platform” most likely comes from itself, “E-commerce”, “SEO-Marketing”, “Social Network App”, “Mobile Device”; the consumption interest in “Entertainment & Trends” most likely comes from itself, “Relationships”, “History & Novels”, and “Culture & Religion.” It reveals that user consumption interest in one topic often comes from the previous consumption interest in similar topics. Figure W6 visualizes the rows of “Zhihu Platform” and “Entertainment and Trends” in Π^{sa} . Interestingly, we find the consumption

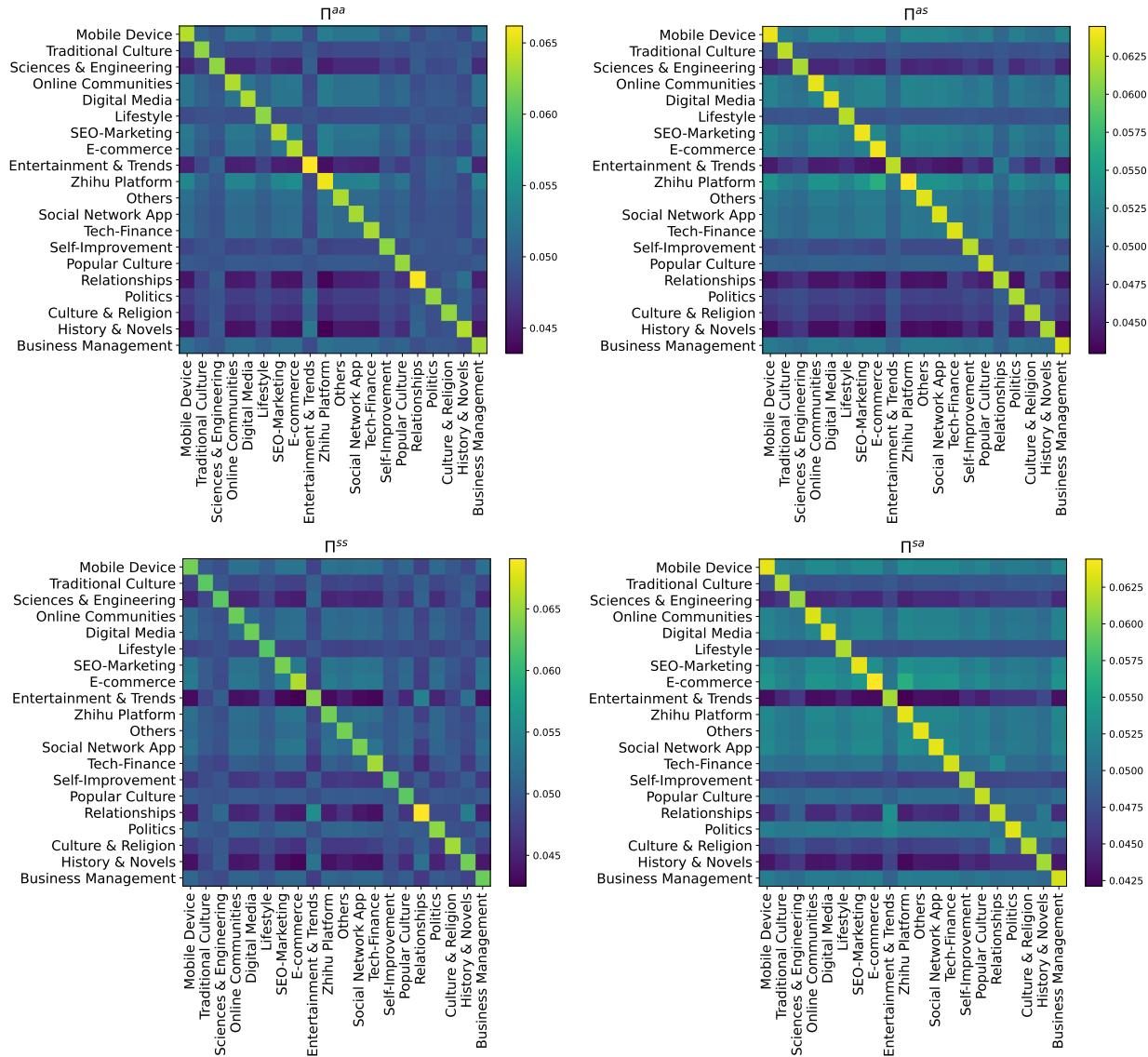


Figure W4: The transition matrices.

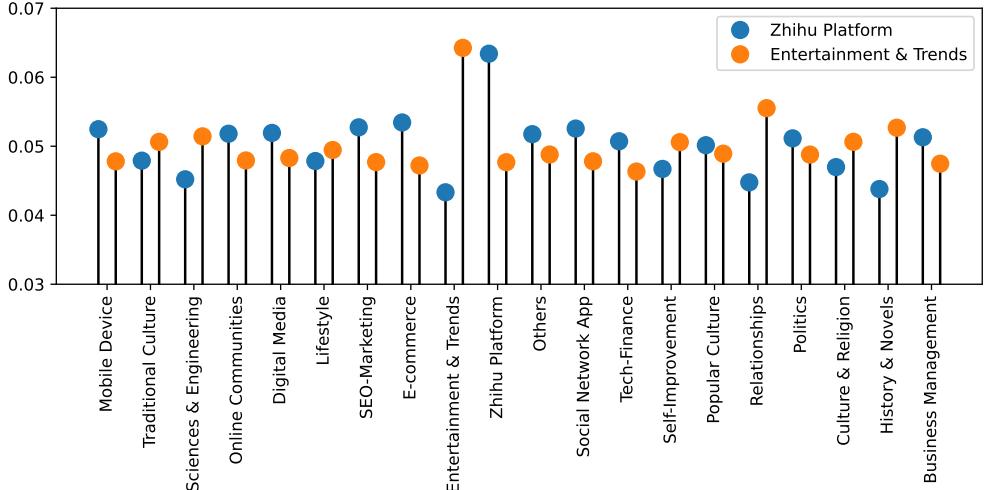


Figure W5: Two Example Topics in the Transition Matrix Π^{ss}

interest in “Entertainment & Trends” mostly comes from the contribution interest in the same topic, “Zhihu Platform”, “Mobile device”, “Digital Media”, “SEO-Marketing.” It suggests that the domain experts in technical fields are drawn to consume less technical topics. This pattern aligns with our findings in the previous subsection that a substantial proportion of experts in hard-oriented fields gradually diversify their contributions to more soft-oriented and general-interest areas.

E-commerce	SEO-Marketing	Social Network App	Mobile device
-3.65%	-3.18%	-3.05%	-3.00%
Relationships	Sciences & Engineering	History & Novels	Entertainment & Trends
3.20%	4.85%	6.36%	6.54%

Table W2: Topics with the Highest and Lowest Growth Rate R_k^{ss}

Table W2 shows the topics with the highest and lowest growth rate R_k^{ss} . We find an interesting shift in consumption interest, transitioning from IT-related topics to general interest topics and STEM subjects. This interest transition aligns with the evolution of Zhihu’s user base, which initially comprised developers and IT professionals and later expanded to the general public, particularly educated young individuals.

C.4 Individual Contribution Preferences

Figure W7 shows the subgroup evolution for the contribution preferences. Table W3 shows the size of each contribution-based cluster at each semi-annual period. Figure W8 shows how the cluster distribution

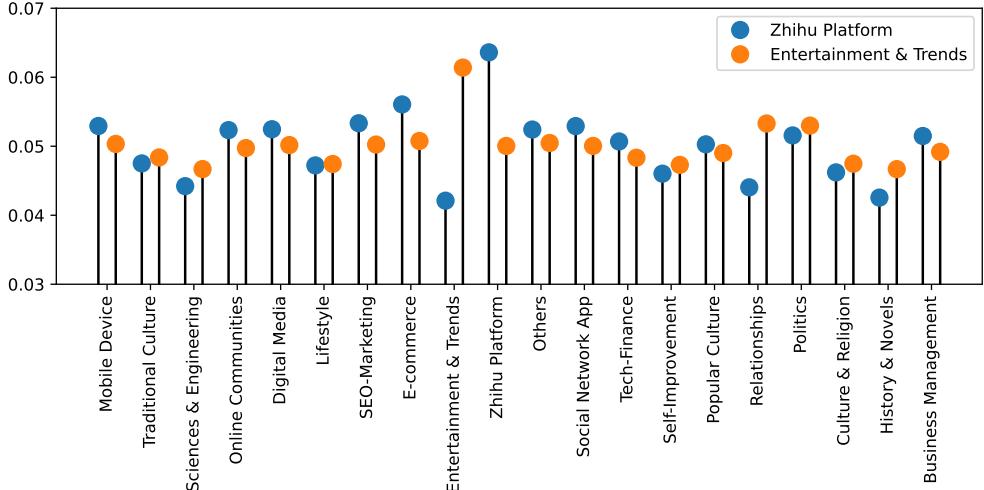


Figure W6: Two Example Topics in the Transition Matrix Π^{sa}

Table W3: Contribution-Based Cluster Size within Each Period

Clusters	1st Period	3rd Period	5th Period
Hard Dabblers	0.430	0.241	0.129
Hard Enthusiasts	0.268	0.290	0.193
Soft Dabblers	0.121	0.187	0.340
Soft Enthusiasts	0.181	0.281	0.340

shifts from the first period to the last period.

C.5 Individual Activities Decomposition

Figure W9 shows the ratio of platform and individual topical strengths for all the topics. Figure W10 and Figure W11 show the dynamic of subgroup-level ratio dynamics for the answering and subscribing activities, respectively.

C.6 Preferences and Follower Size

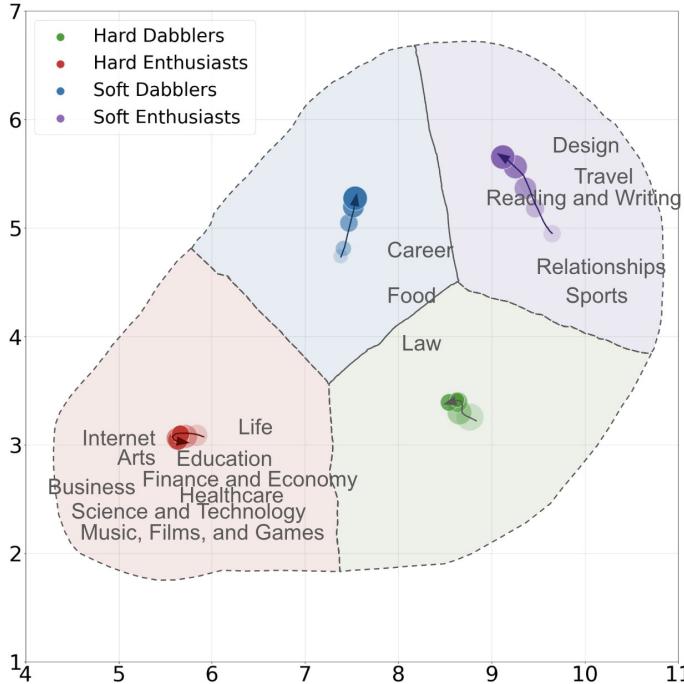


Figure W7: Visualizing the Evolution of Individual Contribution Preferences

Note: The dotted lines mark cluster boundaries, while nodes represent cluster centroids, with size indicating the number of users in each cluster. Lighter node colors correspond to earlier periods, and arrows depict the movement.

Table W4: Relationship Between User Cluster Affiliations and Follower Size.

	DV: $\ln(Follower + 1)$			
	The First Period		The Third Period	
$SoftDabblers_a$	0.792*** (0.081)	0.638*** (0.081)	0.792*** (0.062)	0.695*** (0.062)
$HardDabblers_a$	-0.018 (0.061)	-0.056 (0.061)	0.630 (0.058)	0.513*** (0.060)
$HardEnthusiasts_a$	1.096*** (0.066)	0.830*** (0.068)	2.245 (0.055)	1.940*** (0.060)
$SoftDabblers_q$		1.223*** (0.074)	0.921*** (0.075)	1.182*** (0.060)
$HardDabblers_q$		0.278*** (0.062)	0.214*** (0.062)	0.521*** (0.064)
$HardEnthusiasts_q$		1.027*** (0.071)	0.757*** (0.073)	1.387*** (0.055)
<i>Intercept</i>	3.476*** (0.051)	3.295*** (0.053)	3.160 (0.065)	2.906 (0.040)
Ajd. R-sq.	0.035	0.029	0.049	0.106
			0.050	0.116

Note: The table reports regressions of log-transformed follower size on user contribution-based cluster affiliation at the first and third semiannual periods in the user lifetime. The dependent variable, follower size, is the cumulative follower count by the end of each user's 30th month on the platform. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

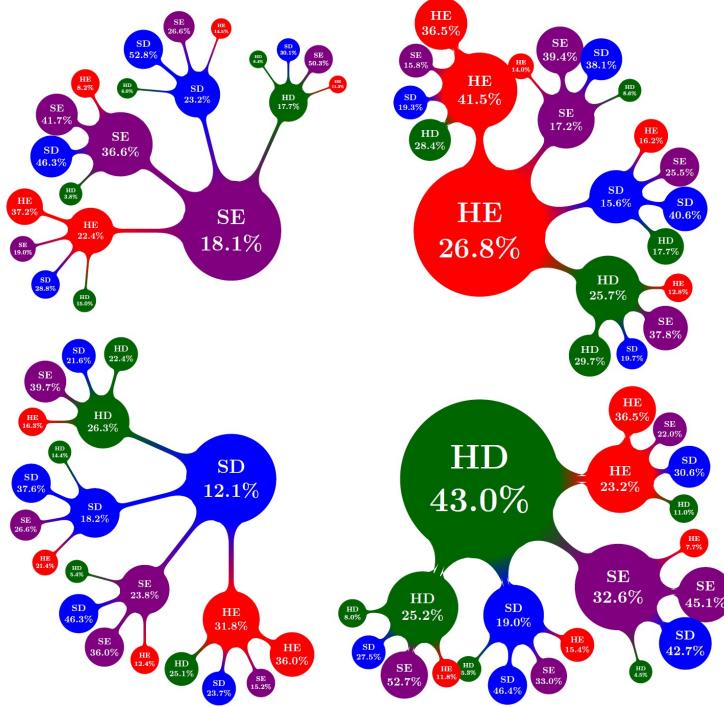


Figure W8: User Migration Patterns across Contribution-Based Clusters

Note: We code each cluster as SE = Soft Enthusiasts, HE = Hard Enthusiasts, SD = Soft Dabblers, HD = Hard Dabblers. The root of each tree represents users belonging to a cluster in the first semiannual period, while the second and third levels correspond to user cluster assignments in the third and fifth semiannual periods, respectively.

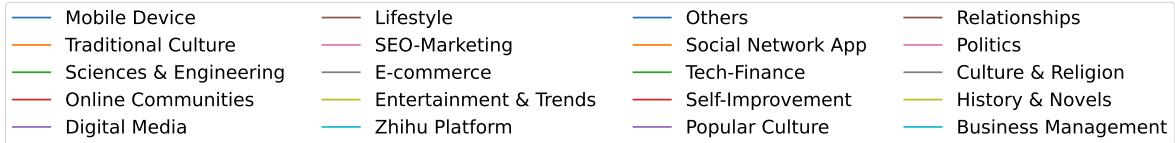
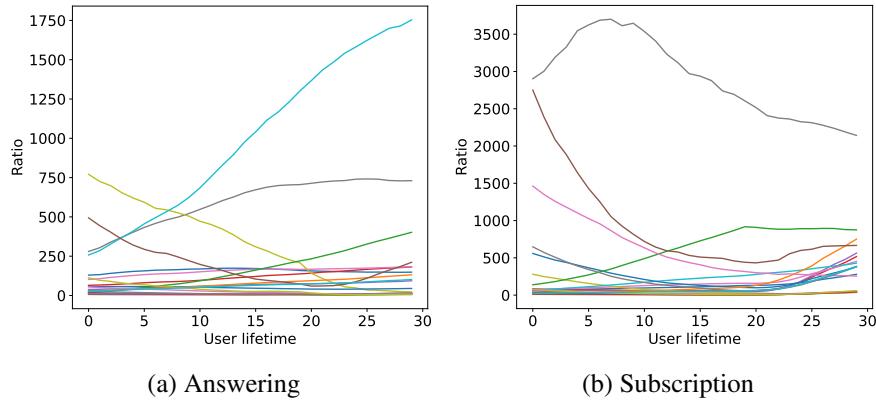


Figure W9: Population Dynamics of Ratio R_{ik} and R_{is}

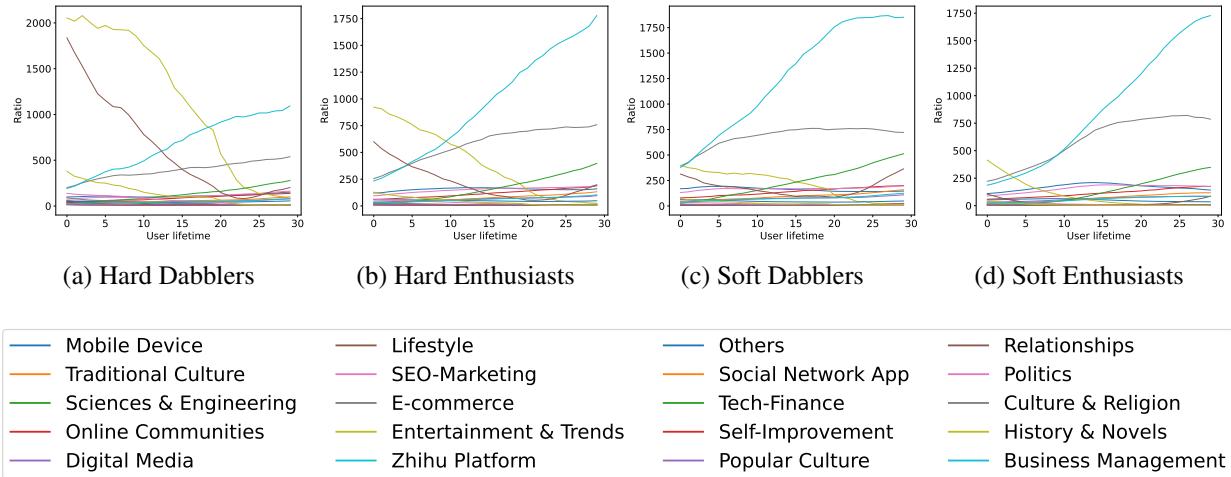


Figure W10: Dynamics of Ratio R_{ik} by Subgroup

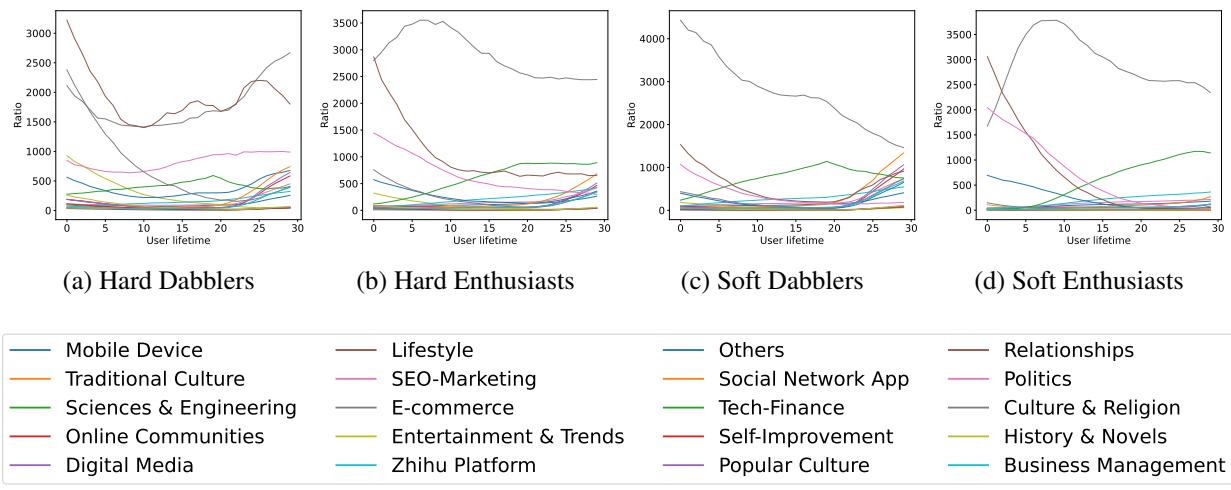


Figure W11: Dynamics of Ratio R_{isk} by Subgroup

References

- Atchison, J. and S. M. Shen (1980). Logistic-normal distributions: Some properties and uses. *Biometrika* 67(2), 261–272.
- Blei, D. M., A. Kucukelbir, and J. D. McAuliffe (2017). Variational inference: A review for statisticians. *Journal of the American statistical Association* 112(518), 859–877.
- Blei, D. M. and J. D. Lafferty (2006). Dynamic topic models. In *Proceedings of the 23rd international conference on Machine learning*, pp. 113–120.
- Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent dirichlet allocation. *Journal of machine Learning research* 3(Jan), 993–1022.
- Burnap, A., J. R. Hauser, and A. Timoshenko (2023). Product aesthetic design: A machine learning augmentation. *Marketing Science*.
- Cain, P. (2022). Modelling short-and long-term marketing effects in the consumer purchase journey. *International Journal of Research in Marketing* 39(1), 96–116.
- Chakraborty, I., M. Kim, and K. Sudhir (2022). Attribute sentiment scoring with online text reviews: Accounting for language structure and missing attributes. *Journal of Marketing Research* 59(3), 600–622.
- Cheng, Z., D. Lee, and P. Tambe (2022). Innovae: Generative ai for understanding patents and innovation. *Available at SSRN* 3868599.
- Dew, R. (2023). Adaptive preference measurement with unstructured data. *SSRN*.
- Dew, R., A. Ansari, and Y. Li (2020). Modeling dynamic heterogeneity using gaussian processes. *Journal of Marketing Research* 57(1), 55–77.
- Dew, R., A. Ansari, and O. Toubia (2022). Letting logos speak: Leveraging multiview representation learning for data-driven branding and logo design. *Marketing Science* 41(2), 401–425.
- Dew, R., N. Padilla, L. E. Luo, S. Oblender, A. Ansari, K. Boughanmi, M. Braun, F. M. Feinberg, J. Liu, T. Otter, L. Tian, Y. Wang, and M. Yin (2024). Probabilistic machine learning: New frontiers for modeling consumers and their choices. *SSRN*.
- Dhillon, P. S. and S. Aral (2021). Modeling dynamic user interests: A neural matrix factorization approach. *Marketing Science* 40(6), 1059–1080.
- Dieng, A. B., F. J. Ruiz, and D. M. Blei (2019). The dynamic embedded topic model. *arXiv preprint arXiv:1907.05545*.
- Dieng, A. B., F. J. Ruiz, and D. M. Blei (2020). Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics* 8, 439–453.
- Duvvuri, S. D., A. Ansari, and S. Gupta (2007). Consumers’ price sensitivities across complementary categories. *Management Science* 53(12), 1933–1945.
- Fox, V. (2007, December). Pew/internet study finds most americans get their answers from the internet. <https://searchengineland.com/pewinternet-study-finds-most-americans-get-their-answers-from-the-internet-13028>.

- Goli, A., P. K. Chintagunta, and S. Sriram (2022). Effects of payment on user engagement in online courses. *Journal of Marketing Research* 59(1), 11–34.
- Gong, C. et al. (2017). Deep dynamic poisson factorization model. *Advances in Neural Information Processing Systems* 30.
- Goodfellow, I., Y. Bengio, and A. Courville (2016). *Deep learning*. MIT press.
- Guadagni, P. M. and J. D. Little (1983). A logit model of brand choice calibrated on scanner data. *Marketing science* 2(3), 203–238.
- Guo, D., B. Chen, H. Zhang, and M. Zhou (2018). Deep poisson gamma dynamical systems. *Advances in Neural Information Processing Systems* 31.
- Hemann, C. and K. Burbary (2013). *Digital marketing analytics: Making sense of consumer data in a digital world*. Pearson Education.
- Hoffman, M. D., D. M. Blei, C. Wang, and J. Paisley (2013). Stochastic variational inference. *The Journal of Machine Learning Research* 14(1), 1303–1347.
- Horrigan, J. B. (2001, October). Online communities. PEW RESEARCH CENTER.
- Jacobs, B., D. Fok, and B. Donkers (2021). Understanding large-scale dynamic purchase behavior. *Marketing Science* 40(5), 844–870.
- Kingma, D. P. and J. Ba (2014). Adam: A method for stochastic optimization. *arXiv*.
- Kingma, D. P. and M. Welling (2013). Auto-encoding variational bayes. *ICLR*.
- Koksal, I. (2020, May). The rise of online learning. <https://www.forbes.com/sites/ilkerkoksal/2020/05/02/the-rise-of-online-learning/?sh=6c9a2daa72f3>.
- Li, H. and L. Ma (2020). Charting the path to purchase using topic models. *Journal of Marketing Research* 57(6), 1019–1036.
- Li, Y., C. Wang, Z. Duan, D. Wang, B. Chen, B. An, and M. Zhou (2022). Alleviating "posterior collapse" in deep topic models via policy gradient. *Advances in Neural Information Processing Systems* 35, 22562–22575.
- Liberali, G. and A. Ferecatu (2022). Morphing for consumer dynamics: Bandits meet hidden markov models. *Marketing Science* 41(4), 769–794.
- Liu, J. and Z. Cong (2023). The daily me versus the daily others: How do recommendation algorithms change user interests? evidence from a knowledge-sharing platform. *Journal of Marketing Research*.
- Liu, J., K. Kawaguchi, and T. Li (2024). Segmenting consumer location-product preferences for assortment localization. *Working Paper*.
- Liu, J., O. Toubia, and S. Hill (2021). Content-based model of web search behavior: An application to tv show search. *Management Science* 67(10), 6378–6398.
- Liu, L. and D. Dzyabura (2021). Capturing heterogeneity among consumers with multitaste preferences. *Available at SSRN 2729468*.

- Liu, X. (2023). Deep learning in marketing: A review and research agenda. *Artificial Intelligence in Marketing* 20, 239–271.
- Liu, X., D. Lee, and K. Srinivasan (2019). Large-scale cross-category analysis of consumer review content on sales conversion leveraging deep learning. *Journal of Marketing Research* 56(6).
- Malhotra, N. K., K. Sudhir, and O. Toubia (2023). *Artificial Intelligence in Marketing*, Volume 20 of *Review of Marketing Research*. Emerald Publishing Limited.
- McInnes, L., J. Healy, N. Saul, and L. Großberger (2018). Umap: Uniform manifold approximation and projection. *Journal of Open Source Software* 3(29), 861.
- Ouyang, S. (2021, January). Zhihu bets big on china's knowledge-sharing market. <https://global.chinadaily.com.cn/a/202101/14/WS5fffde7ca31024ad0baa29fb.html>.
- Rhoades, S. A. (1993). The herfindahl-hirschman index. *Fed. Res. Bull.* 79, 188.
- Ruiz, F. J. R., S. Athey, and D. M. Blei (2020, March). SHOPPER: A probabilistic model of consumer choice with substitutes and complements. *The Annals of Applied Statistics* 14(1), 1–27.
- Schein, A., S. Linderman, M. Zhou, D. Blei, and H. Wallach (2019). Poisson-randomized gamma dynamical systems. *Advances in Neural Information Processing Systems* 32.
- Schein, A., J. Paisley, D. M. Blei, and H. Wallach (2015). Bayesian poisson tensor factorization for inferring multilateral relations from sparse dyadic event counts. In *Proceedings of the 21th ACM SIGKDD International conference on knowledge discovery and data mining*, pp. 1045–1054.
- Schein, A., H. Wallach, and M. Zhou (2016). Poisson-gamma dynamical systems. *Advances in Neural Information Processing Systems* 29.
- Shin, Y., C. Ryo, and J. Park (2014). Automatic extraction of persistent topics from social text streams. *World Wide Web* 17, 1395–1420.
- Sisodia, A., A. Burnap, and V. Kumar (2023). Automatic discovery and generation of visual design characteristics: Application to visual conjoint. Available at SSRN 4151019.
- Sun, M. and F. Zhu (2013). Ad revenue and content commercialization: Evidence from blogs. *Management Science* 59(10), 2314–2331.
- Toubia, O. (2021). A poisson factorization topic model for the study of creative documents (and their summaries). *Journal of Marketing Research* 58(6), 1142–1158.
- Wainwright, M. J., M. I. Jordan, et al. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning* 1(1–2), 1–305.
- Wallach, H. M., I. Murray, R. Salakhutdinov, and D. Mimno (2009). Evaluation methods for topic models. In *Proceedings of the 26th annual international conference on machine learning*, pp. 1105–1112.
- Wang, H. and D.-Y. Yeung (2020). A survey on bayesian deep learning. *ACM computing surveys*.
- West, R., J. Leskovec, and C. Potts (2021). Postmortem memory of public figures in news and social media. *Proceedings of the National Academy of Sciences* 118(38), e2106152118.
- Wu, X., X. Dong, T. T. Nguyen, and A. T. Luu (2023). Effective neural topic modeling with embedding clustering regularization. In *International Conference on Machine Learning*, pp. 37335–37357. PMLR.

- Yang, S. and G. M. Allenby (2003). Modeling interdependent consumer preferences. *Journal of Marketing Research* 40(3), 282–294.
- Yao, E., G. Zheng, O. Jin, S. Bao, K. Chen, Z. Su, and Y. Yu (2014). Probabilistic text modeling with orthogonalized topics. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pp. 907–910.
- Yin, J., Z. Wang, Y. K. Feng, and Y. Liu (2022). Modeling behavioral dynamics in digital content consumption: An attention-based neural point process approach with applications in video games. *Marketing Science, Forthcoming*.
- Yin, M. and M. Zhou (2018). Semi-implicit variational inference. In *International Conference on Machine Learning*, pp. 5660–5669. PMLR.
- Yu, Y., X. Si, C. Hu, and J. Zhang (2019). A review of recurrent neural networks: Lstm cells and network architectures. *Neural computation* 31(7), 1235–1270.