# Meta-Learned User Preference for Topic Participation Prediction

Fan Zhou*, Xiuxiu Qi*, Xovee Xu*, Jiahao Wang*‡, Ting Zhong*, Goce Trajcevski†
*University of Electronic Science and Technology of China, China
†Department of Electrical and Computer Engineering, Iowa State University, USA
‡ Jiahao Wang is the corresponding author: wangjh@uestc.edu.cn

***Abstract*—Predicting the potential user interest on topics in online social networks is important for many practical applications such as advertising, recommendation and malicious account identification. Previous methods on such topic prediction problem mainly focus on learning user preference from historical posting content, and/or rely on the interest of friends to infer the topics a user may be interested in. However, these methods fail to take full advantage of high-order interactions between users and topics and the implicit relations among users, which may result in limited performance. In addition, existing approaches usually require a large amount of samples to train the model and therefore have poor prediction performance for the users who have few content and/or rarely follow the topics. To overcome these limitations, we present a novel method MetaTP (Meta learning based Topic Prediction) for exploiting the complex preference of users over the topics and identify the potential topics for cold-start users. MetaTP is built on a fast graph convolutional network to estimate the user interest through extracting user posting behavior from historical posting content and recursively aggregating the interest from the social friends of a user. Moreover, MetaTP introduces a new way of training prediction model in a meta-learning manner, which not only improves the performance on topic prediction but also can effectively and efficiently adapt to users with a few records. We validate our MetaTP model on real-world datasets crawled from popular social platforms and the empirical results show that our approach significantly outperforms the state-of-the-art baselines.**

***Index Terms*—topic participation, graph neural networks, meta-learning, few-shot learning, social networks.**

## I. Introduction

Nowadays, we are inevitably involved in various online social medias (OSM) such as Twitter, Facebook, Weibo, and Instagram in our daily life. The popular social networks usually have billions of active users, who are sharing experiences, updating photos, commenting on news articles, and are also participating in various discussions or following topics of interest, etc. As for the social platforms, the more users engaging in various topics, the more visibility and business profits they can benefit from advertisements, recommendations, commodity selling and even public health concerns. When we open Twitter or Weibo, the most popular topics (also called hashtags) are usually shown in the obvious positions in the APPs, which could attract more and more people to read, share and comment on those topics. Therefore, predicting which topics users may be interested in has emerged as an important problem in both industry and academia [1]–[3].

Previous efforts on topic prediction/recommendation mainly focus on mining users' posting behaviors and publication contents. For example, earlier works [4]–[6] studied different types of relationships among users and topics through mining user-generated content (UGC) and opinions, and make personalized topic or article recommendations. Tan et al. [7] investigated users' textual reviews and ratings to model users' topics of interest and used the matrix factorization for topic recommendation. Huang et al. [2] try to capture the preference propensity between a user and a few topics, while Liang et al. [8] propose a topic clustering model to track the users' time-varying topic of interests. These methods mainly focus on discovering users' preferences based on UGC (e.g., texts, photos, ratings), which, however, limits their performance for the users with few historical records.

To recommend topics to cold-start users, collaborative filtering (CF), as a well-established technique, is a promising method for addressing the cold start problem. A variety of works [9]–[12] extracted metadata and contextual information from OSM users for tweet recommendation, while topic models (e.g., Latent Dirichlet Allocation) and ranking algorithms (e.g., Bayesian Personalized Ranking) have been widely used for improving the CF-based recommendation performance. Despite achieving promising results on topic recommendation and, to some extent, alleviate the cold-start problem by CF-based correlation learning, these works fail to model the social interactions among users which contain rich information that could affect user behaviors, e.g., people are usually influenced by crowds, especially friends and celebrities, a.k.a. herding behavior in psychology.

Since people are more likely to follow popular and/or emerging topics being extensively discussed in their social circles, leveraging these information, combined with their historical behaviors, would be potentially helpful in learning users' preference and predicting topic participation. Inspired by recent advances in applying deep learning techniques to graph-structured data, we present a graph neural network (GNN) based framework for learning users' preferences. We propose to model the high-order connectivity between users through designing a fast graph convolutional network method to propagate user's preference recursively in social networks.

Instead of relying on a single GNN model that would be

This is the author version of the paper presented at IEEE GLOBECOM '20
doi: 10.1109/GLOBECOM42002.2020.9322197

problematic for cold-start users, we present the Meta-learning based Topic Prediction (**MetaTP**) model to quickly adapt to users with few generated content. MetaTP enables us to estimate users' preferences directly based on an individual user's generated content and her/his social interactions, even when only a small amount of data are available for that user. In contrast to CF-based method that relying on similarity between user posting behaviors, the proposed model generalizes UGC and social relations in a unified framework. More importantly, MetaTP is capable of stacking a number of mini-tasks for enhancing the topic prediction performance while substantially improving the user preference learning in an ensemble learning manner. In particular, we make the following contributions:

- We highlight the critical importance of explicitly exploiting both user-generated contents and social interactions for topic participation prediction and propose a fast graph convolutional network for capturing implicit interactions between users and topics and learning high-order user preference.
- We propose a meta-learning based framework for learning a robust topic participation prediction model. Our model learns prior knowledge from users with rich content and can quickly adapt to users with only limited historical data. To our knowledge, MetaTP is the first attempt for learning user preference by seamlessly integrating graph learning and meta-learning.
- We conduct extensive experiments on the real-world datasets of four popular online social networks including Twitter, Weibo, Zhihu and Douban. The results demonstrate that our MetaTP model significantly outperforms the state-of-the-art methods on all datasets.

## II. Related Work

Understanding users' topics of interest is an important step for many practical applications, ranging from recommendation and advertising, through trending topic and public health concern prediction, to rumor and fake news identification. Many research works have been conducted to discover users' preferences. Most of them [2], [4]–[8] study the UGCs, such as microblogs, articles, photos and videos to uncover user preference over various topics. Typical machine learning techniques such as topic models, classification and clustering methods can be used to predict the topics a user may be interested in. However, these methods pay much attention to historical user contents yet fail to model the interactions between users, which are indicative signals of user future behavior.

Recent advances in deep learning have spurred a variety of graph deep learning methods [13]–[15], which mainly focus on learning representations of graph structures that can be used for many downstream tasks such as node classification and link prediction. Recently, graph neural networks have gained increasing research interests as a means for robust and universal graph representation learning. Generally, GNN models follow a predefined neighborhood aggregation strategy, where the representation of a particular node is learned by recursively aggregating and transforming representations from its neighboring nodes. Various GNN models, such as GCN [15], GAT [16], GIN [17], GraphSAGE [18], etc., have been proposed which vary from each other in the node aggregation mechanisms. As an important application of GNNs, node classification is to identify the labels of nodes, which, however, is slightly different from the topic prediction problem studied in this work. The latter is a multi-label classification problem that cannot be addressed by existing GNN models that mainly focus on multi-class node classification, i.e., each node only belongs to one of the classes. In addition, current GNN models can neither accurately classifying nodes with few data, nor effectively adapt to unseen topics with few participants.

Meta-learning (a.k.a. *learning to learn*) [19], has drawn a significant attention in recent years, due to its capability of quickly adapting to new tasks and leveraging prior knowledge for learning a new concept. The paradigm of meta-learning has been admitted as the most similar way of approximating human intelligence – since humans are naturally able to learn new concepts with learned prior knowledge. Meta-learning methods are usually trained with many mini-tasks and tested on their ability to learn new concepts, which is different from mainstream machine learning techniques. In this work, we borrow the idea of MAML [20] for training our MetaTP, which learns a parameter update scheme that a topic prediction model can take to successfully adapt to the new topics. However, instead of directly applying MAML, we show that slightly modifying the training objective of MAML, we can learn user preference in a traditional supervised learning manner rather than only adapting to new tasks as in MAML. In this vein, we present a new paradigm combining the graph structural learning ability of GNN and the ensemble learning nature of meta-learning for training an enhanced topic prediction model.

## III. Preliminaries

In this section, we formally define the problem and describe the basic features used to learn users' latent topic preferences: (i) content feature – posting history (e.g., user historical tweets); (ii) network feature – the social network structures (e.g., following relationship), as necessary backgrounds of the topic participation prediction problem.

### A. Problem Definition

Our goal is to predict the topics that a particular user would join in the future. Since a user may be interested in more than one topic, we formulate the topic participation (TP) prediction task as a multi-label classification (MLC) problem. Specifically, let $\mathcal{V} = \{v_1, \cdots, v_n\}$ denote $n$ users and given a topic space $\mathcal{Y}$ with $m$ topics: $\mathcal{Y} = \{y_1, \cdots, y_m\}$, i.e., $\mathcal{Y}$ is the vocabulary of all topics discussed by users in $\mathcal{V}$. The TP task is to identify a list of possible topics $\mathbf{y} \in \mathcal{Y}$ for each user $v \in \mathcal{V}$ by training a classification model $f_\theta(\cdot)$. Thereby, the problem can be calculated as finding an optimal topic sequence that maximizes the conditional probability: $p(\mathbf{y}|v) = \prod_{\iota=1}^{l} p\left(y_\iota | y_1, \cdots, y_{\iota-1}, v\right)$, where $l$ is the number of topics that user $v$ participates in.

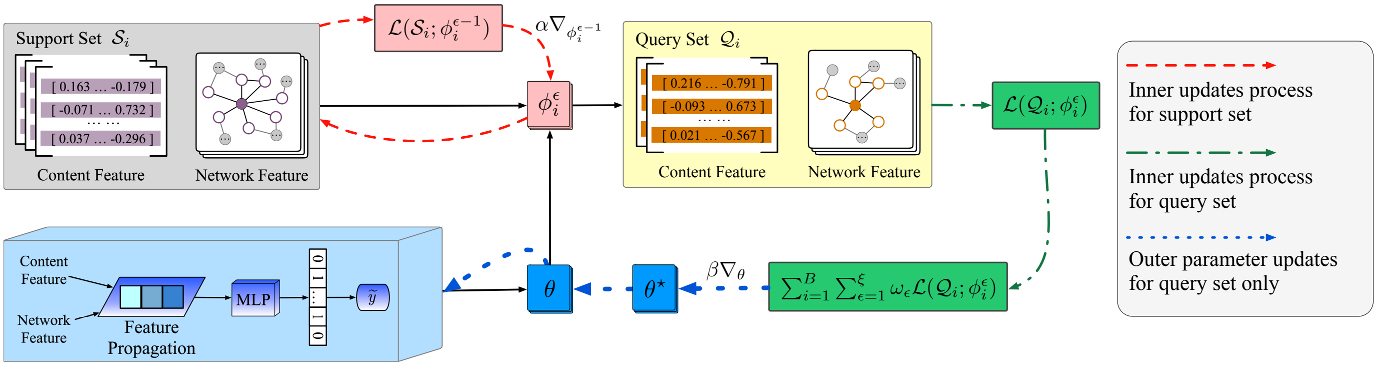



Fig. 1. The framework of MetaTP. It utilizes the content feature and network feature for user preference learning.

### B. Features

We consider two feature types: content and network feature. **Content feature**: for each user $v$, we aggregate all posting history (posting content and user's topic participation history) to form a document and embed the entire document into a vector representation. Here we use doc2vec [21] to obtain a numerical representation that captures the context of the document, denoted as the user content feature matrix $\mathbf{X}$.
**Network feature**: according to previous studies [3], [22], users' behavior is more likely to be influenced by their friends or the crowds on OSM, e.g., users might be interested in topics that most of their friends participate in. We define a graph $G = (\mathcal{V}, \mathbf{E}, \mathcal{N}(v))$ with a symmetric adjacency matrix $\mathbf{E} \in \mathbb{R}^{n\times n}$ and a diagonal degree matrix $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_n)$: (i) $\mathcal{V}$ ($|\mathcal{V}| = n$) is a set of user nodes, and each edge $e_{i,j} = (v_i, v_j) \in \mathbf{E}$ represents an interaction between $v_i$ and $v_j$; (ii) $\mathcal{N}(v)$ denotes the set of $v$'s neighbors (friends) that connect to $v$ within $h$-hops (here $h = 3$); (iii) each entry on the diagonal matrix is equal to the row-sum of the adjacency matrix. Thereby, each node $v$ in the graph has a corresponding $d$-dimensional vector ($d_i = \sum_j e_{ij}$).

## IV. METHODOLOGY

In this section, we describe the details of our proposed meta-learning based topic participation model (MetaTP). Fig. 1 shows the meta-training process of the proposed model, consisting of the task training and parameter updating procedure. In the next, we first present a fast graph convolutional neural networks (GCN) architecture for user interest learning, and then introduce the task (episode) sampling method, which is the basic background of meta-training. Next, we explain how to train the meta-learning tasks to obtain the optimal parameters for improving the prediction performance and adapting to users with few training data.

### A. Learning Aggregated User Preference with GCN

To learn the user preference over topics, we take into account both their historical posting content and social relations. Towards this goal, we present a graph convolutional networks (GCN) [15] based network to learn the representation of user nodes by smoothing and aggregating their content features. It also plays the role of generating a topic list for each user (i.e., predicting the labels). However, directly applying GCN in MetaTP is problematic, because original GCN model requires the nonlinear activation in each aggregation layer, which is computation intensive for our model since we need to train a large number of tasks (will be discussed later). To speed up the training of MetaTP, we introduce a lightweight GCN as the basic network architecture in our MetaTP inspired by [23].

Specifically, we simplify GCN by removing the nonlinearities and alternatively smoothing the hidden feature aggregation with linear transformations. The graph convolution is therefore iteratively computed as:

$$\mathcal{H}^{(\tau+1)} = \mathbf{D}^{-\frac{1}{2}}(\mathbf{A}+\mathbf{I})\mathbf{D}^{-\frac{1}{2}}\mathcal{H}^{(\tau)}, \tag{1}$$

where $\mathcal{H}^{(\tau)}$ denotes the feature representation in the $\tau$-th layer ($\tau \in [0, t]$), $\mathbf{I}$ is an identity matrix added to $\mathbf{A}$ to include self-loops, $\mathbf{D}$ is the degree matrix of $\widetilde{\mathbf{A}} = \mathbf{A}+\mathbf{I}$, and $\mathbf{S} = \mathbf{D}^{-\frac{1}{2}}\widetilde{\mathbf{A}}\mathbf{D}^{-\frac{1}{2}}$ denotes the normalized adjacency matrix with self-connections. In the sequel, the final preference representation of users obtained in the last layer ($t$-th layer) is used for downstream prediction task, expressed as:

$$\begin{aligned}\mathcal{H}^{(t)} &= \mathbf{S}\mathcal{H}^{(t-1)} = \mathbf{S}^t\mathcal{H}^{(0)}\\ &= \tbinom{t}{0}\mathcal{H}^{(0)} + \tbinom{t}{1}\mathbf{S}\mathcal{H}^{(0)} + \tbinom{t}{2}\mathbf{S}^2\mathcal{H}^{(0)} + \cdots + \tbinom{t}{t}\mathbf{S}^t\mathcal{H}^{(0)}\\ &= \tbinom{t}{0}\mathbf{X} + \tbinom{t}{1}\mathbf{S}\mathbf{X} + \tbinom{t}{2}\mathbf{S}^2\mathbf{X} + \cdots + \tbinom{t}{t}\mathbf{S}^t\mathbf{X}.\end{aligned} \tag{2}$$

After getting the final graph representation, we use MLP to output the predicted labels:

$$\tilde{\mathbf{y}} = \mathbf{W}_2\,\mathrm{ReLU}(\mathbf{W}_1\mathcal{H}^{(t)}), \tag{3}$$

where $\mathbf{W}_1$ and $\mathbf{W}_2$ are parameters of MLP which map the feature representation $\mathcal{H}^{(t)}$ to the potential interested topics $\tilde{\mathbf{y}}$ for each user. Here we use ReLU as the nonlinear activation.

### B. Task Sampling

The goal of the meta-learning approaches is to learn a model that can quickly adapt to new tasks with a few data samples by extracting task-general prior knowledge from a number of few-shot sample tasks. Previous meta-learning models for image and text classification follow a episode (task) training paradigm, where tasks $\mathcal{T}$ are typically characterized as $C$-way/class $K$-shot learning, i.e., there are $C$ classes in each task and $K$ training samples in each class.

Formally, we split the training set $\mathcal{D}^{train}$ to sample a set of tasks $\mathcal{T} = \{\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_M\}$ from the task distribution $\mathcal{T} \sim \mathcal{P}(\mathcal{T})$. Meta-learner can learn the common structure in this task distribution, which can be exploited for faster learning with a small number of data available. Each task $\mathcal{T}_i \in \mathcal{T}$ is formed by randomly sampling $C$-topics and $K$ *labeled* users per class as the support set $\mathcal{S}_i = \{(v_{i,j}, y_{i,j})\}_{j=1}^{C\times K}$ and $P$ *unknown* samples from the rest users as the query set $\mathcal{Q}_i = \{(v_{i,k}, y_{i,k})\}_{k=1}^{C\times P}$. The task loss is computed by the error between the model's output $\tilde{\mathbf{y}}_j$ and corresponding ground-truth labels $\mathbf{y}_j$. Since TP task is an MLC problem, we use the cross-entropy loss to calculate the task loss.

During meta-training, we extract $M$ tasks from $\mathcal{P}(\mathcal{T})$ and feed the corresponding data to the GCN model.

### C. Meta Training

The meta-learning algorithm aims to train an effective model that can rapidly adapt to tasks that have not been used during training with few examples [20], [24], [25]. The optimization-based meta-learning algorithm considers the meta-learner $f_\theta(\cdot)$ to learn a good parameters initialization for fast adaptation [25]. In our setting, one may directly utilize meta-learning methods such as MAML [20] for updating the parameters, which provides a gradient-based procedure with only a single additional parameter (the meta-learning rate) and operates on the same parameter space for both meta-learning and fast adaptation. Nevertheless, MAML is a traditional meta-learning method that originally designed for better adapting to the classes that are unseen during meta-training. In contrast, topic prediction requires predicting the topics that a user may participate in, which is a typical multi-label classification problem. In MetaTP, we present a simple yet effective method to modify the MAML for explicitly *learning to learn* from the given support set $\mathcal{S}_i$ to minimize a loss over the query set $\mathcal{Q}_i$, which consists of an inner update and outer update process while making the topics in meta-training and testing the same.

*1) Inner updates:* Formally, we train the MetaTP by optimizing the initial parameters $\theta$ such that after a few-steps $\epsilon = [0, 1, \dots, \xi]$ of gradient descent on $\mathcal{S}_i$, MetaTP can obtain the task-specific parameters $\phi_i$ with acceptable performance on $\mathcal{Q}_i$. We refer to the inner gradient descent procedure that computes $\phi_i$ ($\phi_i^0 = \theta$) as fast adaptation. Since $\phi_i$ is an iteration of a gradient descent procedure that starts from $\theta$, each $\phi_i$ is of the same dimension as $\theta$:

$$\phi_i^\epsilon = \phi_i^{\epsilon-1} - \alpha \frac{\partial \mathcal{L}(\mathcal{S}_i; \phi_i^{\epsilon-1})}{\partial \phi_i^{\epsilon-1}}, \tag{4}$$

where $\alpha$ is the task-learning rate specified as a hyperparameter.

*2) Outer updates:* The ultimate goal of MetaTP is to optimize the model parameters $\theta$ such that a small number of gradient steps on a new task will produce maximally effective topic prediction performance on that task during training. After each inner update, the model parameters $\theta$ in overall meta-objective $\mathcal{L}(\mathcal{Q}_i; \phi_i^l)$ are updated using gradient descent:

$$\begin{aligned}\mathcal{L}(\mathcal{Q}_i; \phi_i^\epsilon) &= \arg\min - \log p(\mathcal{Q}_i | \underbrace{\phi_i^{\epsilon-1} - \alpha \nabla_{\phi_i^{l-1}} \mathcal{L}(\mathcal{S}_i; \phi_i^{\epsilon-1})}_{\phi_i^\epsilon}) \\ &= \mathcal{L}(\mathcal{Q}_i; \phi_i^{\epsilon-1} - \alpha \nabla_{\phi_i^{\epsilon-1}} \mathcal{L}(\mathcal{S}_i; \phi_i^{\epsilon-1})),\end{aligned} \tag{5}$$

where $\mathcal{L}(\cdot)$ denotes the task-dependent cross-entropy loss. The meta-learner is optimized by backpropagating the errors through the task-specific parameters to their common initialization parameters $\theta$. The task loss is defined as:

$$\mathcal{L}(\mathcal{T}_i; \theta) = \sum_{i=1}^{M} \mathcal{L}(\mathcal{Q}_i; \phi_i). \tag{6}$$

Above meta-training method works by minimizing the query set loss computed by base-network after it has completed all inner-updates on a support set per task. Following [26], at every gradient step, the meta-learner's weights were optimized implicitly as a result of backpropagation, which caused many of the instability issues MAML had. To this end, we utilize multi-step loss optimization to improve gradient propagation. Specifically, we use the weighted sum of query set loss after updating every support set loss as the task loss. That is, in a mini-batch $B$, the gradient-based updating rule is:

$$\theta^\star = \theta - \beta \frac{\partial \sum_{i=1}^{B} \sum_{\epsilon=1}^{\xi} \omega_\epsilon \mathcal{L}(\mathcal{Q}_i; \phi_i^\epsilon)}{\partial \theta}, \tag{7}$$

where $\beta$ is the meta-learning rate and the loss weight $\omega_\epsilon$ is the query set loss at $\epsilon$-th step used for improving the gradient stability of meta training. The pseudo-codes for training MetaTP is outlined in Algorithm 1.

Finally, the topic prediction is to produce a list of topics ranking by the output of possibilities output by the MetaTP model, i.e., $\tilde{\mathbf{y}} = f_{\theta^\star}(v^{test})$, where $v^{test}$ denotes the testing users.

## V. Experiments

In this section, we evaluate our model MetaTP against several strong baselines on four real-world datasets.

### A. Experimental Setup

**Datasets.** We prepared four datasets crawled from four major social media platforms, including two microblog platforms `Twitter` and `Weibo`, a Q&A platform `Zhihu` (the largest online Q&A website in China akin to Quora) and a web 2.0 websites `Douban` (provides social networking service for users to share content on topics of movies, books, music, etc.). For all datasets, we preprocess the data as follows: (i) we remove the English stop-words and special characters; (ii) we ignore isolated users in the social network. For each dataset, we build the training and test sets according to the topic participation time. Note that the users in each dataset are ensured to be fully connected after preprocessing. The datasets after preprocessing are summarized in Table II.

**Experimental settings.** For training our MetaTP, we use Adam optimizer for optimization. Although it is common to

**Algorithm 1:** Meta Training Algorithm.

**Input:** training tasks $\mathcal{T}^{train} = (\mathcal{S}^{train}, \mathcal{Q}^{train})$; task-based learning rate $\alpha$, meta-learning rate $\beta$, the number of inner loop updates $\xi$.
**Output:** optimal model parameters $\theta^{\star}$.

Initialize $\theta$ randomly;
**while** not converge **do**
  Sample a batch of tasks $\mathcal{T}_i \in \mathcal{T}^{train}$, $i = [1, B]$;
  /* Outer updates */
  **foreach** $\mathcal{T}_i$ **do**
    $\mathcal{T}_i = \mathcal{S}_i \cup \mathcal{Q}_i$, $\phi_i^0 = \theta$, $\mathcal{L}(\mathcal{T}_i) = 0$;
    /* Inner updates */
    **for** $\epsilon = 1, \cdots, \xi$ **do**
      Compute $\mathcal{L}(\mathcal{S}_i; \phi_i^{(\epsilon-1)})$ and evaluate $\nabla_{\phi_i^{(\epsilon-1)}} \mathcal{L}(\mathcal{S}_i; \phi_i^{(\epsilon-1)})$ using $\mathcal{S}_i$;
      Compute adapted parameters $\phi_i^{\epsilon}$ with gradient descent: $\phi_i^{(\epsilon-1)} - \alpha \nabla_{\phi_i^{(\epsilon-1)}} \mathcal{L}(\mathcal{S}_i; \phi_i^{(\epsilon-1)})$;
      Evaluate weighted loss $\omega_l \mathcal{L}(\mathcal{Q}_i; \phi_i^{\epsilon})$ on query set $\mathcal{Q}_i$;
    **end**
    $\mathcal{L}(\mathcal{T}_i) = \sum_{\epsilon=1}^{\xi} \omega_{\epsilon} \mathcal{L}(\mathcal{Q}_i; \phi_i^{\epsilon})$;
  **end**
  Calculate task loss: $\sum_{i=1}^{B} \sum_{\epsilon=1}^{\xi} \omega_{\epsilon} \mathcal{L}(\mathcal{Q}_i; \phi_i^{\epsilon})$;
  Update $\theta$ by: $\theta^{\star} = \theta - \beta \nabla_{\theta} \sum_{i=1}^{B} \sum_{\epsilon=1}^{\xi} \omega_{\epsilon} \mathcal{L}(\mathcal{Q}_i; \phi_i^{\epsilon};)$.
**end**

use threshold calibration algorithms for multi-label classification, we use the constant 0.5 as the prediction threshold to reduce the impact of external factors. At the meta-level, the values of parameters $C$-way $K$-shot are set to 5-way 5-shot for each meta-training task if not otherwise specified. And the batch size of MetaTP is set to 32. We set the task-learning rate $\alpha$ and meta-learning rate $\beta$ to be 5e-5 and 5e-4, respectively.
**Baselines.** We compare MetaTP to following baseline models: (1) **node2vec** [14] performs the 2nd-order random walks to explore neighborhood architecture and embed nodes with the Skip-Gram model, which is used for user social representation learning in our experiments; (2) and (3) **GCN-F/I** [15]: GCN is a semi-supervised classification method based on graph convolution. Following [3], we derive two topic prediction methods called **GCN-F** and **GCN-I**, respectively using posting embeddings and identity matrix as the input feature matrix in GCN; (4) **Text-associated Deep Walk (TADW)** [27] incorporates text features of vertices into network representation learning under the framework of matrix factorization; (5) **CANE** [28] learns context-aware embeddings for users and models the semantic relationships between users; (6) **LRCNN** [29] presents a CNN architecture for reranking pairs of short texts, and learns the optimal representation of text pairs and a similarity function to relate them in a supervised way; (7) **MACNN** [2] is a CNN-based method for topic prediction, which models user's posting history and topics with external neural memory architecture and attention mechanism.
**Evaluation Metrics.** Following standard metrics of the multi-label classification in topic participation prediction [2], [3], [29], we adopt Accuracy (ACC), Macro-Precision (Macro-P) and Macro-F1 to evaluate the performance of models.

### B. Overall Performance

The results of different models are reported in Table I, where the best results are highlighted in bold. According to this table, MetaTP performs the best in all cases, e.g., the relative improvement over the best baseline range from 6% to 22% in terms of Macro-P across four datasets, showing that our model is more effective on learning user preference. In addition to overall topic prediction performance, we have following observations.

First, previous network representation learning models are effective for topic participation prediction, even without considering user-generated contents, which can be proved by the fact that node2vec sometimes performs best among the baselines. This interesting finding suggests that user content and social relations are not consistent for discovering users' real preferences, e.g., users' preference on topics (e.g., on `Zhihu` and `Twitter`) may change significantly with time, which means their historical contents are not necessary and are even controversial to their future participated topics. This argument can be further proved by the results that two text mining based models LRCNN and MACNN are not comparable, even compared with simply network embedding techniques. These observations necessitate rethinking our traditional ways of user preference learning in such a period that the user's favor changes rapidly, and also proposes an interesting question, i.e., how to capture the changes of user preference.

Second, simply relying on attributed graph learning (e.g., TADW, CANE, and GCN) may results in worse prediction performance. Surprisingly, we found GCN is not sensitive to user-generated content (e.g., the performances of GCN-I and GCN-F are very close), which verifies the motivation of this work, i.e., there are implicit interactions between user and topics, which can not be captured by a single graph convolutional network. In contrast, our method adapts to the users with few samples by learning prior knowledge of user preference from a (large) number of tasks and is therefore more robust to perturbations of user preference changes.

Finally, the prediction performances of various models, vary differently on different platforms. On `Weibo` dataset, the prediction task is relatively easy since there are fewer topics for each user. However, all models perform poorly on `Zhihu` and `Twitter`, due to more topics users may participate in. Surprisingly, models achieve better prediction results although there are significantly more topics on `Douban`. The rationale behind this phenomenon is that `Douban` is a website for users to share personal topics of interest which is relatively stable, compared to the trending topics in `Zhihu` and `Twitter`. Though there are more topics in `Douban`, the users are closely clustered by the topics (or groups) which makes the prediction simpler than predicting the rapidly changed trending topics.

### C. Parameter Sensitivity.

We conduct a sensitivity analysis to investigate the influence of two important parameters of MetaTP. Due to the lack of

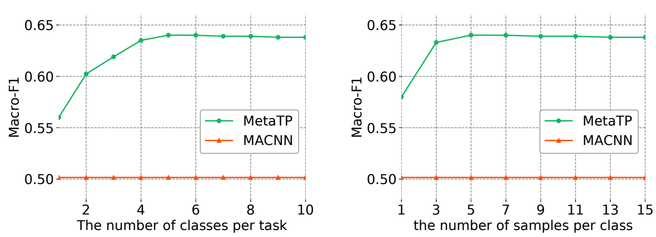
TABLE I
PERFORMANCES COMPARISONS ON FOUR DATASETS.

| Method | Weibo | | | Zhihu | | | Twitter | | | Douban | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Macro-P | Macro-F1 | ACC | Macro-P | Macro-F1 | ACC | Macro-P | Macro-F1 | ACC | Macro-P | Macro-F1 | ACC |
| node2vec | 0.4238 | 0.3803 | 0.3480 | 0.1062 | 0.0597 | 0.0387 | 0.2690 | 0.1297 | 0.0793 | 0.4478 | 0.3418 | 0.2336 |
| GCN-I | 0.1750 | 0.2196 | 0.1534 | 0.0711 | 0.1137 | 0.0631 | 0.1538 | 0.2127 | 0.1239 | 0.3811 | 0.3694 | 0.2405 |
| GCN-F | 0.2000 | 0.2725 | 0.1812 | 0.0707 | 0.1155 | 0.0637 | 0.1544 | 0.2229 | 0.1306 | 0.3848 | 0.3727 | 0.2450 |
| TADW | 0.4739 | 0.4208 | 0.3870 | 0.0176 | 0.0070 | 0.0047 | 0.2394 | 0.1270 | 0.0805 | 0.4530 | 0.1582 | 0.0959 |
| CANE | 0.2832 | 0.2529 | 0.2299 | 0.0399 | 0.0204 | 0.0138 | 0.2125 | 0.0981 | 0.0586 | 0.4454 | 0.3390 | 0.2449 |
| LRCNN | 0.3550 | 0.3200 | 0.2923 | 0.1061 | 0.0700 | 0.0441 | 0.1994 | 0.1520 | 0.0952 | 0.4472 | 0.3269 | 0.2261 |
| MACNN | 0.5640 | 0.5016 | 0.4645 | 0.0767 | 0.0245 | 0.0155 | 0.2670 | 0.1129 | 0.0724 | 0.4451 | 0.3782 | 0.2768 |
| **MetaTP** | **0.6813** | **0.6401** | **0.5811** | **0.1261** | **0.1163** | **0.0726** | **0.3272** | **0.2616** | **0.1731** | **0.4790** | **0.4118** | **0.3045** |

TABLE II
BASIC STATISTICS OF WEIBO, ZHIHU, TWITTER, DOUBAN.

| | Weibo | Zhihu | Twitter | Douban |
|---|---|---|---|---|
| #Topics | 12 | 89 | 55 | 154 |
| #AvgTopics[a] | 1.5 | 6.1 | 8.5 | 56.7 |
| #Users | 6,732 | 2,368 | 2,673 | 2,241 |
| #Tweets | 84,168 | 754,015 | 828,254 | 1,362,789 |
| #Friends | 96,496 | 374,925 | 106,244 | 145,962 |
| #Tasks | 10,195 | 3,552 | 4,000 | 3,300 |
| Training data | 2013.01.14-2013.12.31 | 2015.06.01-2016.12.31 | 2015.01.01-2016.12.31 | 2016.01.01-2016.12.31 |
| Testing data | 2014.01.01-2014.05.12 | 2017.01.01-2017.12.22 | 2017.01.01-2017.12.31 | 2017.01.01-2017.12.31 |

[a] denotes the average number of topics a user participates in.




(a) Effect of $C$. (b) Effect of $K$.

Fig. 2. The impact of parameters on Weibo.

space, we only report the results on Weibo dataset, but note that the results are similar on other datasets. Fig. 2 illustrates the impact of parameters $C$ and $K$ on Macro-F1 score.

*a) Effect of $C$:* $C$ determines the number of classes per task. We can observe that the performance of MetaTP initially increases with the value of $C$ but soon becomes stable. This is reasonable since the tasks are sufficiently sampled, a few classes in each task are enough to train the MetaTP.

*b) Effect of $K$:* we then examine the influence of $K$, i.e., the number of samples for each class in the support set. The best performance of MetaTP is achieved when $K = 5$. This result indicates that our model can adapt to predicting users' topics of interests with significantly fewer samples.

## VI. CONCLUSION

In this paper, we proposed a novel GNN-based meta-learning paradigm for learning online social users' preference for predicting the topics a user would participate in. Specifically, the proposed method MetaTP unifies user-generated content and social network influence in a graph convolutional network by aggregating features from high-order connectivities. Our model can not only identifies users' personalized preferences, but also being able to estimate user preferences with only a small number of contents through learning prior knowledge to adapt to tasks. Empirical results demonstrate that the proposed model outperforms baselines on different social networks.

## ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (Grant No.61602097), and NSF grants CNS 1646107.

## REFERENCES


[1] A. Gough, R. F. Hunter, O. Ajao, A. Jurek, G. McKeown, J. Hong, E. Barrett, M. Ferguson, G. McElwee, and M. McCarthy, "Tweet for behavior change: using social media for the dissemination of public health messages," *JMIR Public Health and Surveillance*, 2017.

[2] H. Huang, Q. Zhang, J. Wu, and X. Huang, "Predicting which topics you will join in the future on social media," in *SIGIR*, 2017.

[3] F. Zhou, L. Liu, K. Zhang, G. Trajcevski, and J. Wu, "Predicting topic participation by jointly learning user intrinsic and extrinsic preference," *IEEE Access*, vol. 7, pp. 8917–8930, 2019.

[4] H. Liang, Y. Xu, D. Tjondronegoro, and P. Christen, "Time-aware topic recommendation based on micro-blogs," in *CIKM*, 2012.

[5] S. Gerani, M. J. Carman, and F. Crestani, "Aggregation methods for proximity-based opinion retrieval," *Acm Transactions on Information Systems*, vol. 30, no. 4, pp. 26.1–26.36, 2012.

[6] A. Alsaudi, M. Sadri, Y. Altowim, and S. Mehrotra, "Adaptive topic follow-up on twitter," in *ICDE*, 2017.

[7] Y. Tan, M. Zhang, Y. Liu, and S. Ma, "Rating-boosted latent topics - understanding users and items with ratings and reviews," in *IJCAI*, 2016.

[8] S. Liang, Z. Ren, Y. Zhao, J. Ma, E. Yilmaz, and M. de Rijke, "Inferring dynamic user interests in streams of short texts for user clustering," *ACM TOIS*, vol. 36, no. 1, pp. 1–37, 2017.

[9] K. Chen, T. Chen, G. Zheng, O. Jin, E. Yao, and Y. Yu, "Collaborative personalized tweet recommendation," in *SIGIR*, 2012.

[10] R. Yan, M. Lapata, and X. Li, "Tweet recommendation with graph co-ranking," in *ACL*, 2012.

[11] X. Cheng, X. Yan, Y. Lan, and J. Guo, "Btm: Topic modeling over short texts," in *TKDE*, 2014.

[12] D. Kim, J.-G. Lee, and B. S. Lee, "Topical influence modeling via topic-level interests and interactions on social curation services," in *ICDE*, 2016.

[13] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in *SIGKDD*, 2014.

[14] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *SIGKDD*, 2016.

[15] T. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *ICLR*, 2017.

[16] P. Velickovic, G. Cucurull, A. Casanova, A. Remoro, P. Liò, and Y. Bengio, "Graph attention networks," in *ICLR*, 2018.

[17] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, “How powerful are graph neural networks?” in *ICLR*, 2018.
[18] W. L. Hamilton, R. Ying, and J. Leskovec, “Inductive representation learning on large graphs,” in *NIPS*, 2017.
[19] S. Thrun and L. Y. Pratt, “Learning to learn,” *Springer Berlin*, vol. 55, no. 3, pp. 193–197, 1998.
[20] F. Chelsea, A. Pieter, and L. Sergey, “Model-agnostic meta-learning for fast adaptation of deep networks,” in *ICML*, 2017, pp. 1126–1135.
[21] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” in *NIPS*, 2013, pp. 3111–3119.
[22] M. Baddeley, “Herding, social influence and economic decision-making: Socio-psychological and neuroscientific analyses,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 365, 2010.
[23] F. Wu, T. Zhang, A. H. d. Souza Jr, C. Fifty, T. Yu, and K. Q. Weinberger, “Simplifying graph convolutional networks,” in *ICML*, 2019.
[24] R. Vilalta and Y. Drissi, “A perspective view and survey of meta-learning,” *Artificial Intelligence Review*, vol. 18, no. 2, 2001.
[25] S. Ravi and H. Larochelle, “Optimization as a model for few-shot learning,” in *ICLR*, 2017.
[26] A. Antoniou, H. Edwards, and A. Storkey, “How to train your maml,” in *ICLR*, 2018.
[27] C. Yang, Z. Liu, D. Zhao, M. Sun, and E. Chang, “Network representation learning with rich text information,” in *IJCAI*, 2015.
[28] C. Tu, H. Liu, Z. Liu, and M. Sun, “Cane: Context-aware network embedding for relation modeling,” in *ACL*, 2017.
[29] A. Severyn and A. Moschitti, “Learning to rank short text pairs with convolutional deep neural networks,” in *SIGIR*, 2015.