Predicting Users' Future Interests on Social Networks: A Reference Framework

Fattane Zarrinkalam[a,*], Havva Alizadeh Noughabi[b], Zeinab Noorian[c], Hossein Fani[d] and Ebrahim Bagheri[e]

[a]*School of Engineering, University of Guelph, Guelph, Canada*
[b]*Department of Computer Engineering, Ferdowsi University of Mashhad, Mashhad, Iran*
[c]*Ted Rogers School of Information Management, Toronto Metropolitan University, Toronto, Canada*
[d]*School of Computer Science, Faculty of Science, University of Windsor, Windsor, Canada*
[e]*Laboratory for Systems, Software and Semantics (LS3), Toronto Metropolitan University, Toronto, Canada*

ARTICLE INFO

Keywords:
Future Interest Prediction
Learning to Rank
Twitter

ABSTRACT

Predicting users' interests on social networks is gaining attention due to its potential to cater customized information and services to the end users. Although previous works have extensively explored how users' interests can be modeled on social networks, there has been limited investigation into the prediction of users' future interests. The objective of our work in this paper is to empirically study the effectiveness of different sets of features based on users' past social interactions, historical interests and their temporal dynamics to predict their interests over a collection of future-yet-unobserved topics. More specifically, we introduce and formalize the features for interest prediction in four categories: *user-based*, *topical*, *explicit user-topic engagement*, and *friends' influence*. We further explore the influence of temporality by augmenting features with information pertaining to users' historical interests and social connections. We model the task of future interest prediction as a learning-to-rank problem where different features and their related categories are ranked based on their relevance and performance in interest prediction, and investigate the efficiency of different features individually and comparatively for predicting the future interest of users with different activity levels in social networks over on unobserved topics. After conducting experiments on a real-world dataset sourced from Twitter, we have identified several noteworthy findings: 1) `relevance` feature in the category of past explicit user-topic engagement is the strongest indicator for predicting user's future interest across all user groups, with an observed 8.57% decrease in NDCG and an 8.95% decrease in MAP when it is removed in the ablation study. 2) the observation of an 8.06% decrease in NDCG and a 7.3% decrease in MAP, when topical features such as `popularity`, `freshness`, and `coherence` are removed in the ablation study, highlights their significance as among the strongest indicators for users' future interest, particularly for low-activity users. 3) although temporal features show a clear positive impact across user groups with varying levels of activity (resulting in a 4.5% decrease in NDCG and a 7.3% decrease in MAP when removed in the ablation study), the temporal topical features do not demonstrate a significant positive effect, and 4) The removal of user-specific characteristics such as `influence` and `personality traits` in the ablation study reveals their significant impact in predicting future interest over *cold* topics, reflected by a 5.49% decrease in NDCG and a 5.72% decrease in MAP. Our findings make significant contributions to the field of future interest prediction, offering valuable insights and practical implications for various applications in social network analysis.

1. Introduction

The extensive volume of user-generated data present on social networks provides a valuable resource for acquiring comprehensive insights into diverse facets of user behavior and preferences (Kursuncu et al., 2019; Salminen et al., 2022; Aldous et al., 2023). Accurately predicting the interests of users improves downstream tasks like personalized advertising and efficient service delivery to customers. It enables decision-makers to strategically anticipate user reactions to potential future topics, facilitating effective future planning. Although predicting the future interests of users on social networks holds significant importance, the existing body of work primarily concentrates on extracting

ORCID(s):

users' present or past interests. Consequently, only a limited number of researches in the literature have been specifically focused on predicting the future interests of users on social networks (Bao et al., 2013; Kang et al., 2019). Future interest prediction aims to determine a user's interest in a set of future topics by examining their historical interests, interactions, and behavior. The limitation of existing works on future interest prediction is the assumption that the set of topics in the future will be the same as the set of topics observed in the past, falling short when predicting the interests of users with regard to future *unobserved* topics. Taking into account the dynamic nature of trending topics on social networks, which can dynamically shift in response to real-world events (Pereira et al., 2018; Abel et al., 2011a), topics of the future might not necessarily be the same as those in the past or present. Thus, it is important to formalize the problem of future interest prediction based on unobserved topics by considering the temporal variability of both user interests and future unobserved topics.

The aim of our research in this paper is to fully understand the landscape of future interest prediction through a comprehensive study of the literature on interest prediction, identifying various forms of features that have been used in the past for similar tasks (not necessarily future interest prediction but broadly interest modeling and detection), and perform an extensive empirical analysis on 19 different feature types and evaluating their individual and collective impact on future interest prediction for future-yet-unobserved topics. We systematically define four categories of features including (1) *user-based features* to derive users' characteristics based on their activities on social networks, (2) *topical features* to consider the characteristics of a topic emerged on social networks, (3) *explicit user-topic engagement features* and (4) *friends' influence features* to investigate the influence of friends on the user's interests. In addition, we define temporal variants for each category of features to consider the temporality of user activities on social networks. In order to operationalize the features in practice, we adopt a learning-to-rank framework to investigate the impact of various combinations of these features for predicting the future interest of users with different levels of activities over unobserved topics.

This paper examines the effectiveness of various feature types for accurately predicting users' future interests on social networks. Prior to conducting the experiment, to examine the effectiveness of the activity level of users on the future interest prediction problem, we categorize users into three groups based on their level of activity: 1) *low-activity* users, 2) *semi-active* users, and 3) *highly-active* users. First, we investigate the predictive value of various user characteristics. Our findings consistently demonstrate the positive impact of *user-based features*, thereby highlighting their significance for accurately predicting users' future interests. Secondly, we explore the effect of *topical features* on future interest prediction. We discover that topical features are strong indicators for predicting the future interests of users with lower levels of activity. Furthermore, we study the relationship between *users' historical interests* and their future interests. By examining the continuity or shifts in user interests, we assess the strength of historical interests as predictors of their future preferences. We also examine the extent to which the historical interests of *friends influence* a user's future interests. Our findings indicate that the future interests of highly active users are significantly influenced by their social connections, emphasizing the pivotal role of social relationships in shaping individual interests. Moreover, we explore the potential improvement in future interest prediction accuracy by incorporating the temporality of user activities. By considering the dynamic nature of user engagement, we find that *temporal* features have a positive impact across user groups with varying levels of activity. Lastly, we aim to identify effective indicators for predicting the users' future interests on *cold* topics. Our results reveal that user-specific characteristics are among the most effective features for predicting future interest in these cases. Our work provides the following specific contributions:

- We introduce and systematically formalize features that enable user modeling by proposing specific quantifiable measures. These measures are subsequently incorporated into a learning-to-rank framework for predicting future unobserved topics of interest;
- We empirically evaluate the introduced features within the context of future interest prediction on future-yet-unobserved topics and their relevance and impact are compared and critically evaluated under different conditions;
- The experimental results not only evaluate the performance of the features, both individually and collectively but also offer insights into the factors influencing the varying performance of specific feature categories under different conditions in predicting future user interests. These insights significantly deepen our understanding

of the underlying mechanisms and contribute to elucidating why certain feature categories exhibit superior (or inferior) performance in the prediction task. These findings contribute significantly to the advancement of user interest prediction methods, offering practical guidance for feature selection and model development in this domain.

The remaining sections of this paper are structured as follows. In Section 2, we provide a review of the related literature covering related works on future interest modeling and prediction. In Section 3, we formulate the problem of future interest prediction. In Section 4, we first present six research questions that guide our investigation, and then we introduce and formalize a set of features specifically tailored for predicting future user interests. Section 5 outlines our experimental setup, dataset, evaluation results, and the analysis of our features. Moving forward to Section 6, we engage in a comprehensive discussion of our research findings based on the aforementioned analysis. Finally, the paper concludes with Section 7.

2. Related Work

While there have been very few studies that have considered the task of predicting users' future interests, there have been other works that have explored user interest modeling. For the sake of being comprehensive, we cover a broad range of works in this space, which can be classified according to the types of information sources they utilize, namely: 1) user-generated textual contents, such as tweets (*content-based*); 2) social network structure that depicts the connections between users as well as their content (*network-based*); and, 3) factors related to time that capture the dynamic aspects in interests of users (*temporal*). In this section, we examine the notable works within each of these categories.

2.1. Content-based Approaches

The primary information source utilized for inferring users' interests from social networks is the textual content generated by users, also known as *social posts*, shared by a user with a possible engagement of other users through, e.g., liking or re-sharing. Social posts of a user often contain a set of keywords that imply the user's topics of interest. In contrast to utilizing low-level features like tokens, certain studies have chosen to derive high-level features like semantic concepts or named entities from the users' posts. These extracted features are then regarded as representative of the topics of user's interest (Zhao et al., 2015). It is common to find hashtags or embedded links (URLs) in social posts to specify their topical matters. Consequently, hashtags and URLs can be recognized as valuable sources of information that aid in identifying a user's interests (Penas et al., 2013; Piao and Breslin, 2016c). For example, Abel et al. (2011b) semantically enriched social posts of users by the content of the news articles whose URLs were mentioned within those posts. Likewise, in Kapanipathi et al. (2011) by annotating users' tweets with DBPedia concepts, their interests have been effectively modeled. Karatay and Karagoz (2015) introduced a Named Entity Recognition (NER) based user profile model to generate personalized interest recommendations. In a like manner, Jipmo et al. (2017) have represented the users' posts and their interests as bags of articles and categories of Wikipedia respectively, and then ranked interests by relevance according to the graph distance between the articles and the categories. Piao and Breslin (2016a) proposed a user modeling strategy that integrates entity and category-based user profiles and showed synergistic improvements compared to category-based as well as entity-based user profiles.

As users' topics of interest are often semantically related to each other, content-based studies have also focused on modeling user interest at different granularity levels: primitive, hierarchical, and implicit interests, based on the relations between topics. For example, Kapanipathi et al. (2014) modeled users' interests at primitive and hierarchical levels. They have undertaken the task of modeling a user's fundamental interests by annotating the posts of users with Wikipedia entities. They then employ a frequency-based scoring mechanism to calculate the user's level of interest in each entity. To extract the hierarchical interests of users, they take into account the relationships between categories in the knowledge base, utilizing the user's primitive interest scores and the hierarchical structure of Wikipedia categories. Piao and Breslin (2016b) leveraged graph-based knowledge to extract the implicit users' interests by incorporating various forms of relationships between entities and categories. This approach allowed them to expand upon the user's primitive interests and deduce their implicit interests more comprehensively. In (Wang et al., 2019), the authors analyze the semantic relationship of user tags and obtain the inter and intra correlations between each tag pair. To infer implicit user interests, they acquire the semantic correlation matrix by combining the above two correlations and further update the user-tags matrix (which represents the user interest) accordingly.

Recently, Qi et al. (2022a) infer user interests from interactions between topics of news articles. The key idea is that capturing the word-level interaction across the news clicked by the same user contains rich clues to understand user interests. Considering inter-news and intra-news word-level interaction and concatenating them into a long document, they utilize a transformer network to efficiently model users' interest. Qiu et al. (2022) introduced three interrelated modules that are able to jointly model both the current and future interests of users. The key idea is to learn the existing interests of users by analyzing their history of interaction with the news recommender system and then leverage knowledge graphs to enrich user representations with their potential interests. This proposed bidirectional interaction layer is able to dynamically build a bipartite graph of user interests and their associated entities to fuse two interest representation models. Qi et al. (2021) represented user models through a hierarchical interest tree to obtain different aspects and coarse-grained and fine-grained user interests. They exploited attention networks across a three-level hierarchy to infer the importance of different topics and subtopics for each user. In similar research (Qi et al., 2022b), Qi et al. proposed a self-attention network that uses the historical activities of users along with their candidate topic of interest (i.e., the to-be recommended topic) to model the global interest of users. Okuda et al. (2023) have presented a novel approach for learning representations, which they refer to as user portraits. These user portraits represent users' diverse interests for recommendation purposes. They have constructed the user portraits by leveraging both the users' utterances and a knowledge graph including information on item classes.

Another information that has been used as an indicator of user interest modeling is users' *personality traits*. The work of Dhelim et al. (2020) integrates the users' personality into their users' interest inference model. They propose a personality-aware system for interest mining that takes into account the personality traits (McCrae and John, 1992) and a dynamic topic model for interest prediction. They have utilized a heterogeneous graph to represent relationships between users (based on their personality analysis), topics and between users and topics and further use meta-path discovery to detect implicit and explicit interest of users. Similarly, Alrehili et al. (2022) combines personality traits and demographic information into user interest modeling. Developing the heterogeneous graphs of users and topics, they utilized the relation between topics as well as demographic and personality similarity between users to predict implicit users' interests.

2.2. Network-based Approaches

Users are often connected to each other through their social connections, which could be used for user interest detection. The theory of Homophily, which describes the users' tendency to form connections with others who share similar preferences or interests (McPherson et al., 2001), is the inspiration for most of the works in this category. Mislove et al. (2010) have utilized it to make inferences about missing information and user interests by leveraging the data shared by the neighbors of users. Pennacchiotti et al. (2012) have obtained the user's interests by examining not only the user's tweets but also the tweets of her neighboring users. In Wang et al. (2014), the argument is that algorithms solely rely on explicit links between users lack effectiveness in the context of sparse and dynamic social networks. To address this limitation, they have introduced a novel approach for extracting user interests by leveraging a specific link structure assumption. This assumption posits that local link structures between nodes can serve as an indication of their similarity. For instance, if two users have a substantial number of common followers, it suggests that they are likely to share similar topical interests. Welch et al. (2011) have already reported that the retweeting relation is more closely correlated with the topical similarity of user interests. By comparing the relationship between retweeting and followership, they have determined that retweeting is a considerably more powerful indicator of topical interest. Furthermore, Wang et al. (2013) have studied the relatedness between the common interests of the users and their social connections like followership, retweet, mention, and comment. In their study, it was observed that followership and retweet relations serve as stronger indicators of connections among users with common interests, in contrast to mention and comment relations. However, their approach is better suited for inferring interests of users from inactive or new users. Xu and Lu (2015) have introduced a unique bi-relational graph model that captures the interactions between users and their shared areas of interest. This model consists of two sub-graphs representing users and topics, enabling us to leverage user homophily and topic correlation concurrently. Through extensive large-scale studies, they gained valuable insights into the effectiveness of inferring user interests by utilizing the underlying social connections.

Social influence, which suggests that a user's interests can be influenced by her friend's interests, is also followed by some other works to utilize relationships between users. Jamali and Ester (2010) have proposed the SocialMF model, which incorporates the impact of users' friendships on their interests. Bao et al. (2013) have adopted the SocialMF model and proposed a temporal and social probabilistic matrix factorization model (TS-PMF) in microblogging services for predicting the users' interests. Budak et al. (2014) have introduced a probabilistic model based on user and network information to extract the latent users' interests from Twitter. They have included the user's susceptibility to the influence of her friends within their model. Bhattacharya et al. (2014) have inferred a user's interests by examining the topical expertise of famous users on Twitter. The methodology is rooted in the observation that users typically follow influential users within the space of their topical interests. Because their approach struggles with non-famous users, He et al. (2015) proposed a modified method for topic modeling on Twitter which relies on information from the underlying relationship network to identify interest tags for non-famous users. In (Bhattacharya et al., 2014) and (He et al., 2015), the researchers extracted the topical expertise of popular users by utilizing the features of their Twitter Lists. Chen et al. (2013) have presented a model that takes users' social influences from their friends into account to estimate the rating that a user will assign to an item at a given time. Their proposed approach comprises two primary steps: first they model user receptiveness at a single time point by extending Bi-LDA to consider social relations in its modeling and in the second step, they incorporate temporal information to allow for modeling over time. Spasojevic et al. (2014) have proposed Large Scale Topic Assignment (LASTA), to mine topical interests of users from multiple social networks. LASTA generates various distinct features by leveraging signals like user textual content and social graph connections. The authors demonstrated that incorporating this diverse range of features results in a more comprehensive representation for topical interests of users, surpassing the utilization of solely generated text or graph-based features.

Social influence is still one of the key factors considered in more recent studies for users' interest modeling in social networks. For example, Deng et al. (2018) presented a topic interest mining algorithm relying on tags and users' bidirectional interaction, focusing on the probability of forward and backward tag distribution in the user's networks. Liu et al. (2023) have proposed a novel social influence model called SIGA for predicting ratings. SIGA combines a graph autoencoder (GAE) to effectively capture user-item interactions in a bipartite graph, while also quantifying social influence based on information dissemination in social networks. This hybrid approach aims to leverage the strengths of both social modeling and GAE, resulting in high-quality representations of users and items for accurate rating prediction. Similarly, another study Yu et al. (2023) utilized an extended variation of the expectation confirmation model to investigate the impact of friendship factors on intention to social commerce continuance. Notably, the findings from this study underscore the particular significance of informational social influence, especially during the early stages of pre-consumption. In another study, Perifanis et al. (2023) introduced FedPOIRec, a solution that preserves privacy while generating recommendations for Point of Interest (POI). By utilizing social relations and combining individual-level parameters over encrypted data, FedPOIRec enhanced recommendation quality and personalization.

2.3. Temporal Approaches

Dynamicity, also known as temporality, of users' interests is a significant factor in the interest prediction task. The rationale behind this is that modeling the user's current interests without regarding the dynamic nature of their interests over time may overlook the evolution of topics and user interests over time. Furthermore, within social networks, both the user's interests and the topics themselves undergo dynamic changes in response to real-world events. An early research, Orlandi et al. (2012) applied a frequency-based approach to compute the degree of user's interest over entities mentioned in the user's textual content. They further considered the temporality in the interests of users by incorporating an exponential time decay function. This function assigns greater significance to more recent occurrences of each entity, thereby assigning a higher weight to recent interests. A temporal probabilistic framework is introduced in (Sang et al., 2015) for determining user interest profiles. This research makes the assumption that users exhibit short-term and long-term interest distributions. In fact, long-term interests represent stable user preferences that persist over time. In contrast, short-term interests reflect the preferences of users regarding temporary topics or events. Bao et al. (2013) introduced a probabilistic matrix factorization model, which combines both temporal and social aspects. This model utilizes sequential interest matrices of users captured at varying time points and incorporates users' friendship matrices. The primary objective of the model is to predict users' future interests by considering both temporal patterns and social influences. Chen et al. (2013) have proposed a temporal model, which focuses on predicting future ratings given by a user, taking into account how users are affected by their friends. Zarrinkalam et al. (2019) presented a temporal content-based recommendation model that employs the Wikipedia category hierarchy. The hierarchy serves as a comprehensive and adaptable topic space, enabling the system to predict user interests for a range of topics that have not been previously observed.

Additionally, Zheng et al. (2016) have introduced an approach based on multi-Markov chains to predict future interests. Their proposed approach demonstrates a high level of accuracy in predicting both long-term and short-term user interests. To deal with the dynamic trend in users' interests, Zheng et al. (2019) exploited the timeliness and interactivity characteristic of microblogging platforms. They developed a three-layer interest network (TIN-LDA) model according to the dynamic interest hierarchical orientation for mining users' interests. The work in (Cheng et al., 2023) leveraged bidirectional GRU and a time adjustment function to detect the interest evolution of users by investigating their interest trajectory data. Specifically, the proposed method is able to account for alternating categories of interest with multi-head attention to improving user interest modeling. Shao et al. (2022) have introduced a framework aimed at investigating and leveraging fine-grained evolving interests to enhance friend recommendation in social networks. Their approach includes extracting interest tags from users and employing an improved TSA-LSTM (Time Series Analysis LSTM) model. This model incorporates two controllers, namely a time-aware controller and a semantic-aware controller, that can capture the evolving users' interests and infer a wider range of precise interest tags.

While previous studies have examined various features for modeling user interests, a systematic analysis of their predictive performance in predicting future interests has mainly remained unexplored. To address this research gap, we undertake an empirical investigation to assess the impact of different features related to user interest modeling for predicting future interests across unobserved topics by categorizing these features into four groups: topical features, user-specific features, explicit topic-user engagement features, and friends' influence features. We further explore the influence of temporality by augmenting features with information pertaining to users' historical interests and social connections. By providing a comprehensive evaluation of both content-based and social connection-based features, while incorporating temporal information, this study significantly contributes to the field by facilitating the development of more accurate prediction models.

3. Problem Definition

The task of predicting users' future interests relies on the temporal dynamics of topics and the historical interests of users. In order to account for temporality, it is necessary to partition the historical activities of users into T discrete time intervals $1 \leq t \leq T$. Within each time interval, it is important to consider a specific set of topics and analyze the distribution of user interests among those topics.

Given $\mathbb{Z}^t = \{z_1, z_2, \ldots, z_K\}$ a set of K topics extracted from published posts in time interval t for each user $u \in \mathbb{U}$, her interest profile at time t over $\mathbb{Z}^t$, denoted by $P^t(u)$, can be defined as:

Definition 3.1 (Interest Profile). The interest profile of user $u \in \mathbb{U}$ in time interval t, relative to a set of topics $\mathbb{Z}^t$, denoted by $P^t(u)$, is expressed as a vector of weights assigned to the K topics, i.e., $(f_u(z_1^t), \ldots, f_u(z_K^t))$, where $f_u(z_k^t)$ represents the extent of u's interest in topic $z_k^t \in \mathbb{Z}^t$. Normalization of the user interest profile is performed using the L1-norm.

It should be noted that topic and user interest detection methods for computing $\mathbb{Z}^t$ and $P^t(u)$ from social networks have been extensively researched and analyzed in existing literature (Zarrinkalam et al., 2018; Trikha et al., 2018), which is why our work does not concentrate on them. Considering $\mathbb{M}^t$, the set of tweets posted in time interval t, a topic modeling method such as Twitter-LDA (Zhao et al., 2011) results in (1) K topic-term distributions ϕ where each topic entity distribution associated with a topic $z \in \mathbb{Z}^t$ represents active topics in $\mathbb{M}^t$, and (2) $|\mathbb{U}|$ user-topic distributions each of which is associated with a user u and represents the interest profile of user u in time interval t, i.e., $P_t(u) = (f_u(z_1^t), \ldots, f_u(z_K^t))$. A topic z can be modeled using top-n term representation, i.e., $z = \{w_1, w_2, ..., w_n\}$ that are ranked based on the term probabilities $p(w|z)$ in the topic-term distribution ϕ.

For a corpus of posts $\mathbb{M}$, and T time intervals, the following two artifacts will need to be produced: (1) A collection set of topics, denoted as $\mathbb{Z}$, which includes T topic sets, i.e., $\mathbb{Z}^1, \mathbb{Z}^2, \ldots, \mathbb{Z}^T$, where $\mathbb{Z}^t$ is a set of K topics extracted from published posts in time interval t, i.e. $\mathbb{M}^t$; (2) A set of users' interest profiles, i.e. $\{P^t(u) | 1 \le t \le T, u \in \mathbb{U}\}$, which include T interest profiles for every user $u \in \mathbb{U}$, i.e., $P^1(u), ..., P^T(u)$.

Considering historical interests in each time interval $t: 1 \le t \le T$, each user will have T user interest profiles, a method for predicting future user interests would need to operate in the following manner:

Definition 3.2 (Future Interest Profile). Given the set of users' interest profiles $\{P^t(u) | 1 \le t \le T, u \in \mathbb{U}\}$ and potential future topics, i.e., $\mathbb{Z}^{T+1}$, the future interest prediction task aims at predicting the future interest profile of each user $u \in \mathbb{U}$ with respect to $\mathbb{Z}^{T+1}$, i.e., $P^{T+1}(u)$.

Based on this task formulation, given a user u, we seek to rank the potential topics in the future, $\mathbb{Z}^{T+1}$, according to their relevance and importance to user u. To this end, we aim to systematically classify and empirically evaluate the impact of potential features that can be effectively used for predicting users' future interests.

4. Research Framework

In this section, we will initially outline the six research questions addressed in this paper. Following that, we will provide a detailed explanation of the introduced features. This formalization comprises four primary categories, namely user-based features, topical features, explicit user-topic engagement features, and friends' influence features.

4.1. Research Questions

Our primary goal is to introduce, systematically classify and empirically study the relevance and impact of different features that can be effective predictors of users' future interests over unobserved topics. To this end, we define a set of six research questions that will be methodically addressed and answered throughout this paper:

RQ1. Which characteristics of *users* on social networks are effective predictors for their future interests?
RQ2. Which characteristics of *topics* have an impact on predicting the future interest of users?
RQ3. Are *historical* interests of a user strong indicators for her future interests?
RQ4. Are future interests of a user influenced by the historical interests of their *friends*? If so, what is the extent of social influence on user's future interests?
RQ5. Does considering the *temporality* of user activities improve the user's future interest prediction?
RQ6. What are effective indicators for predicting the future interest of users on *cold topics*?

4.2. Feature Formalization

This section presents the formalization of the features that may be influential in predicting the future topical interest of users on social networks. Given the features defined in the existing literature related to user interest modeling and our research questions, we have categorized possible relevant features from four main perspectives: (1) *user-based* features in which we consider the characteristics of the user such as her `influence`, `prestige` and `personality traits`; (2) *topical* features considering different topic-related features such as `popularity`, `freshness`, `coherency` and `exclusivity`; (3) *explicit user-topic engagement* features where we consider the relatedness of user's historical content to the potential future topics such as user's `content relevancy` and user's `interest coherence`; (4) *friends' influence* features that consider the influence of a user's friends on the user's interests such as `user susceptibility` and relational interests. In addition, we consider *temporal* features as a cross-cutting feature where we add a time dimension to the aforementioned features to investigate their impact on the accuracy of the future interest prediction problem. We introduce various features in each category by proposing concrete quantifiable metrics, which are further systematically evaluated for their relevance and impact on the future interest prediction task in Section 5. Figure 1 depicts the hierarchical organization of these features in a tree-like structure. Table 1 summarizes the four distinct feature categories that might be impactful on future interest prediction.

4.2.1. Topical Features

To model a user's interests over a given topic on social networks, topic-based indicators consider the characteristics of that topic. These features are quantifiers for the topic's importance and have the potential to provide insights to predict the interest's degree of users to a topic.

Topic Popularity: Previous research shows the topic's `popularity` is one of the most impactful features for user interest modeling (Mathioudakis and Koudas, 2010; Song et al., 2012; Yin et al., 2015). We conjecture that a higher topic `popularity` increases the likelihood of users being interested in that topic. Thus, given the viral nature of information dissemination on social networks, we define the `popularity` of a topic by the number of users who are interested in that topic as follows:

$$\texttt{popularity}(z) = \frac{|\mathbb{U}_z|}{|\mathbb{U}|} \tag{1}$$

where $\mathbb{U}_z$ denotes the set of users who are interested in z and $\mathbb{U}$ denotes all the users.

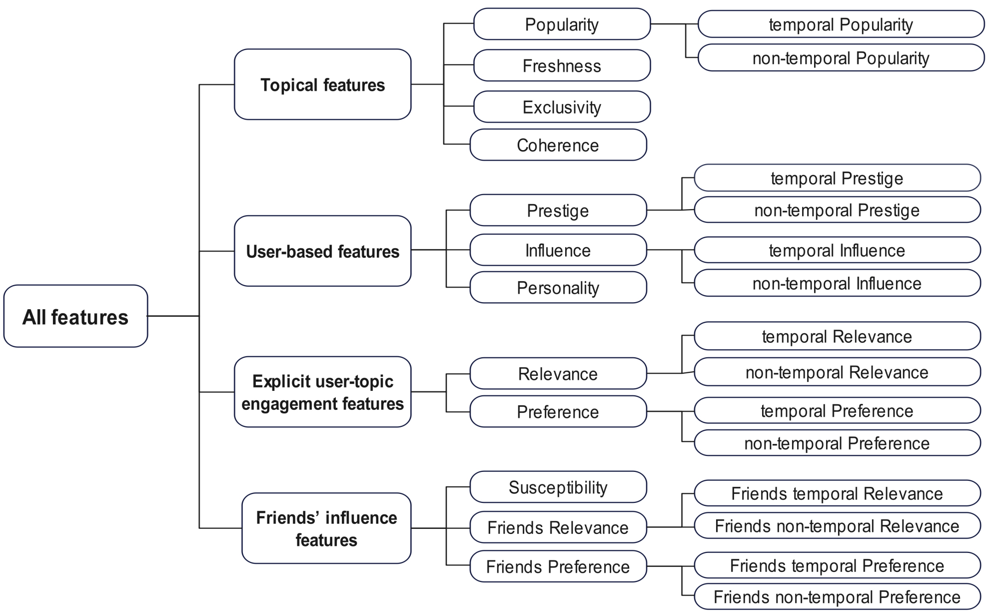

Figure 1: Illustration of all features in a tree structure.

Table 1
Summary of features introduced in our paper.

Category	Feature Name	Description	Citations
Topical features	Popularity	The ratio of users who are interested in topic z to the total number of users.	(Mathioudakis and Koudas, 2010) (Song et al., 2012) (Yin et al., 2015)
	Temporal Popularity	The ratio of users who are interested in topic z to the total number of users across different time intervals.	
	Freshness	How much recent topic z has emerged.	(Kacem et al., 2014) (Das et al., 2015) (On-At et al., 2017)
	Exclusivity	How often the top words of topic z do not appear as top words in other topics.	(Bischof and Airoldi, 2012) (Fani et al., 2018)
	Coherence	How much the top words in topic z tend to co-occur together.	(Fang et al., 2016) (Saraswat and Chakraverty, 2022)
User-based features	Prestige	The number of prestigious users has retweeted the posts of user u.	(Musiał et al., 2009) (Yang et al., 2012) (Rong and Mei, 2013) (Freire et al., 2022)
	Temporal Prestige	The number of prestigious users have retweeted the posts of user u across different time intervals.	
	Influence	How much attention user u has received from other users by retweeting her posts.	(Cha et al., 2010) (Kanavos and Livieris, 2020) (Trigka et al., 2022)
	Temporal Influence	How much attention user u has received from other users by retweeting her posts across different time intervals.	
	Personality	How much of each personality trait (openness to experience, conscientiousness, agreeableness, extraversion and neuroticism) is present in the textual content of user u.	(Ribeiro et al., 2018) (Dhelim et al., 2020) (Lampropoulos et al., 2022) (Alrehili et al., 2022)
Explicit user-topic engagement features	Relevance	How much topic z is related to the textual content of user u.	(Abel et al., 2011a) (Zarrinkalam et al., 2015) (Inaba and Takahashi, 2018) (Qiu et al., 2022)
	Temporal Relevance	How much topic z is related to the textual content of user u across different time intervals.	
	Preference	How much topic z is similar to the u's topics of interest.	(Wen and Lin, 2011) (Shen et al., 2013) (Bhattacharya et al., 2014) (Zarrinkalam et al., 2018) (Trikha et al., 2018)
	Temporal Preference	How much topic z is similar to the u's topics of interest across different time intervals.	
Friends' influence features	Susceptibility	The number of terms that user u has mentioned in her posts by getting influenced from what her friends have previously published.	(Budak et al., 2014) (Lee and Lim, 2015) (Hassan et al., 2016) (Galal et al., 2021)
	Friends Relevance	How much topic z is related to the textual content of u's friends.	(Chen et al., 2013) (Bao et al., 2013) (Spasojevic et al., 2014) (Budak et al., 2014) (Wang et al., 2014) (Wang et al., 2018) (Liu et al., 2023) (Perifanis et al., 2023)
	Friends Temporal Relevance	How much topic z is related to the textual content of u's friends across different time intervals.	
	Friends Preference	How much topic z is similar to the topics of interest of u's friends.	
	Friends Temporal Preference	How much topic z is similar to the topics of interest of u's friends across different time intervals.	

Topic Freshness: Topic recency or the `freshness` of a topic (i.e., whether the topic has recently gained significant attention in discussions) has been considered as another feature in predicting whether the topic will gain users' attention (Das et al., 2015; Kacem et al., 2014; On-At et al., 2017). The key intuition is that the topics that have emerged recently are more likely to receive a higher degree of attention in the future. Therefore, we measure a topic's `freshness` based on the time that it has emerged, as follows:

$$\texttt{freshness}(z) = 1 - \frac{L - t_z}{L} \tag{2}$$

where t_z denotes the time interval in which the topic z has emerged for the first time and L is the number of historical time intervals.

Topic Exclusivity: The intuition behind topic `exclusivity` is that when a new topic that has not been previously observed, emerges on social networks, it has the tendency to attract the attention of a wider range of users. Topic `exclusivity` has been shown to be impactful in studying user modeling on social networks (Bischof and Airoldi, 2012; Fani et al., 2018). Thus, in order to investigate whether exclusive topics will gain more attention from users in the future, we measure the exclusivity of a topic based on the extent to which the top terms associated with the topic are not found among top terms in other topics, which means the less the terms observed in a given topic are, the more exclusive that topic will be (Bischof and Airoldi, 2012). Formally:

$$\texttt{exclusivity}(z) = \frac{1}{|z|} \sum_{w \in z} \frac{1}{|\mathbb{Z}_w|} \tag{3}$$

where $\mathbb{Z}_w$ includes the set of topics which include the term w in their top-10 words.

Topic Coherence: Topic `coherence`, alongside topic popularity, can affect the likelihood of users developing an interest in a particular topic. Topic `coherence` is a metric to measure the interpretability and meaningfulness of topics (Fang et al., 2016; Saraswat and Chakraverty, 2022). In (Fang et al., 2016), the authors proposed several topic `coherence` metrics to automatically evaluate topic quality. They have shown that metrics based on Pointwise Mutual Information are the most similar to human judgment. Therefore, we adopt this metric to calculate topic `coherence`:

$$\texttt{coherence}(z) = \frac{1}{|X|} \sum_{(w_i, w_j) \in X} PMI(w_i, w_j) \tag{4}$$

where $X = \{(w_i, w_j) | w_i, w_j \in z \ and \ i \neq j\}$ and $PMI(w_i, w_j)$ is pointwise mutual information (PMI) score of word pairs of topic z that appear in each post.

4.2.2. User-based Features

Previous research in user modeling on social networks signifies the importance of incorporating users' engagement and analyzing the language employed in their posts in user characterization (Alavijeh et al., 2023). Thus, in this subsection, we intend to derive users' characteristics from various perspectives to evaluate whether they can serve as an indicator in predicting future users' interests.

User Prestige: Prestigious users within social networks are followed/retweeted by many other users; their perceived prestige is further elevated when their followers themselves possess a prestigious status (Yang et al., 2012; Musiał et al., 2009; Rong and Mei, 2013; Freire et al., 2022). Users with high `prestige` have the power to shape other's interests because other users on social networks find it desirable to follow the interests of these users (Musiał et al., 2009). Hence, we assume that considering historical users' `prestige` can provide insight into predicting users' future interests. Given a directed graph, a robust measurement to estimate the `prestige` of a node is PageRank (Brin, 1998), based on which the node `prestige` depends on the respective `prestige` of the nodes linking to them (Rong and Mei, 2013; Gayo-Avello, 2013). We measure the `prestige` of a user on social networks based on the retweet graph of users as follows:

$$\texttt{prestige}(u) = PageRank(u, rtG) \tag{5}$$

where $rtG = (\mathbb{V}, \mathbb{E}, g)$ is a retweet graph which is a weighted directed graph that nodes $\mathbb{V}$ represents users $\mathbb{U}$ and edges $\mathbb{E}$ are formed by observing any action of retweet relation between two users. If a user v has retweeted at least one of u's posts, there is an edge from u to v. Further, function g calculates the weight of each edge on the graph based on the number of times that u's posts are retweeted by v.

User Influence: A user is considered to be influential on social networks if her actions affect the actions of many other users. A user's `influence` is usually due to her outstanding fame or expertise in a specific domain (Cha et al., 2010; Riquelme and González-Cantergiani, 2016; Kanavos and Livieris, 2020; Trigka et al., 2022). As a result, an influential user is able to provoke social activity and interest and act as a hub for attracting, directing and boosting the interests of other users in those domains. Despite user prestige which is primarily associated with the structural position of users within a network, the influence of a user is not dependent on the influence status of the users who interact with them. Thus, to measure user `influence` on social networks, we estimate the ratio of attention the user has received from other users by retweeting her posts on social networks, to the number of posts she has posted (Cha et al., 2010). Formally:

$$\texttt{influence}(u) = \frac{1}{|I_u|} \sum_{v \in I_u} (|RT_{vu}|/|RT_v|) \tag{6}$$

where I_u is the set of followers of the user u, RT_{vu} denotes the u's posts which are retweeted by user v and RT_v is the set of posts retweeted by user v.

Personality Traits: Previous studies have shown the importance of `personality traits` in characterization of users in social networks (Gil de Zúñiga et al., 2017; Dhelim et al., 2020; Yin et al., 2020; Salminen et al., 2020; Lampropoulos et al., 2022). For instance, `personality traits` of Twitter users have been analyzed to understand hateful behavior and to classify users accordingly (Ribeiro et al., 2018). One of the most popular methods used for personality detection is the Five Factor Model (aka Big Five), introduced by Tupes and Christal (Tupes and Christal, 1992). The Big Five model hypothesizes that human psychological characteristics can be summarized in five aspects, namely, *p1: openness to experience, p2: conscientiousness, p3: agreeableness, p4: extraversion, p5: neuroticism*. To investigate whether the future interest of users is affected by their `personality traits`, we adopt the Big Five model and represent each user by a vector over these five dimensions of personality, denoted by $P(u) = (q_u(p_1), .., q_u(p_4), q_u(p_5))$, where $q_u(p_i)$ is a function that calculates relatedness value between the tweets of user u and the personality trait p_i. calculates the mean value of the tweets associated with each user, which are labeled by personality p_i.

4.2.3. *Explicit User-topic Engagement Features*

We consider the historical content of the user as a potential indicator to predict her future interests, which also has been extensively examined in the user modeling literature (Abel et al., 2011a; Zarrinkalam et al., 2015; Inaba and Takahashi, 2018; Di Tommaso et al., 2018; Bennacer Seghouani et al., 2019). We elaborate this feature under two concepts: `users' content relevancy` and `users' interest coherence`, as follows.

User's Content Relevancy: Inspired by earlier user interest modeling studies (Abel et al., 2011a; Zarrinkalam et al., 2015; Inaba and Takahashi, 2018), we hypothesize that the relevancy of a user's generated content to a target topic in the future is an indicator for her future interests. Therefore, given a user-topic pair (u, z), we define the `relevance` function, which measures the relatedness of a potential topic in the future z to the textual content of user u as follows:

$$\texttt{relevance}(u, z) = \frac{\sum_{w \in z} tf(w, W_u)}{\sum_{z' \in \mathbb{Z}, z' \neq z} \sum_{w \in z'} tf(w, W_u)} \tag{7}$$

where $tf(w, W_u)$ denotes the term frequency of the topic term $w \in z$ in the textual content of user u, i.e.W_u.

User's Interest Coherence: Several studies utilize semantic relationships between topics on social networks to measure the degree of interest a user has in a particular topic. These studies mostly rely on the assumption that users often have coherent and related interests (Wen and Lin, 2011; Shen et al., 2013; Bhattacharya et al., 2014; Zarrinkalam et al., 2018; Trikha et al., 2018). Inspired by this, we conjecture that the users are likely to show interest in topics that are conceptually similar to the topics they have previously been interested in. To formulate this, we define the `preference` function, which measures the similarity of a potential topic in the future z to the historical topics of interest of the user u, i.e., $\mathbb{Z}_u$. Formally:

$$\texttt{preference}(u, z) = \frac{1}{|\mathbb{Z}_u|} \sum_{z' \in \mathbb{Z}_u} sim(z, z') \tag{8}$$

where $\mathbb{Z}_u$ represents the topics of interest of the user u (i.e. the set of topics in which the interest's degree of the user u is more than zero, formally $\mathbb{Z}_u = \{z \in \mathbb{Z} | \ \exists t \ f_u^t(z) > 0, \ 1 \leq t \leq T\}$). Further, $sim(z, z')$ calculates the cosine similarity between two topics z and z'.

4.2.4. Friends' Influence Features

Existing literature has considered the impact of social influence on shaping users' interest profiles on social networks. Here, we consider some of these features as follows:

User Susceptibility: one of the well-studied features that impacts user interest modeling is referred to as user `susceptibility` (Budak et al., 2014; Lee and Lim, 2015; Hassan et al., 2016; Galal et al., 2021). The idea is that what users publish on social networks is influenced by not only what their friends have published but also susceptible to how the users are influenced by their friends. This is also supported by social influence theory indicating that a user's behavior and opinions are affected by the behaviors of her friends. To examine the impact of considering this feature on the user interest prediction problem, we need to define a `susceptibility` function. As proposed in (Budak et al., 2014), we calculate a user's `susceptibility` at a given time interval t based on the posts that her friends have published in the previous time interval $t-1$. That is, the more the user uses the terms mentioned in the posts of her friends, the more susceptible the user is. Since the `susceptibility` of a particular user can differ from one friend to another, the `susceptibility` of a user is calculated by taking into account the extent to which the user is susceptible to being influenced by each of their friends. Afterward, we combine these values by using an average aggregation function. Formally:

$$\texttt{susceptibility}(u) = \frac{1}{|F_u|} \sum_{v \in F_u} sus(u, v) \tag{9}$$

where $sus(u, v)$ is measured based on the number of times that the user u uses the terms mentioned by her friends in the previous time interval. As previously indicated in Section 3, in order to account for the temporal dynamics of topics and users' historical interests, we partition the users' past activities into T distinct time intervals, where $1 \leq t \leq T$. Therefore, the definition of $sus(u, v)$ is as follows:

$$sus(u, v) = \frac{1}{|T-1|} \sum_{t=2}^{T} \frac{|W_u^t \cap W_v^{t-1}|}{|W_u^t|} \tag{10}$$

where W_u^t denotes the terms that user u has mentioned in her posts in time interval t.

User's Relational Interest: It is also possible to measure the interest degree of a user in a given topic by taking into account the users' social connections. Based upon social influence theory, certain studies have taken social connections into account–relying on the premise that users tend to influence and hence have the likelihood of exhibiting similar interests (Chen et al., 2013; Bao et al., 2013; Spasojevic et al., 2014; Budak et al., 2014; Wang et al., 2014, 2018). In the context of future interest prediction, based on the fact that the more a user's friends engage with a topic, the more likely it will be for the user to be interested in that topic in the future, we formalize this feature as follows:

Given a pair of user-topic (u, z), we define a function to calculate the influence of u's friends on her potential interest in topic z. To do so, we define two functions, i.e. $\texttt{friendsInfluence}_{\texttt{Relevance}}(u, z)$ and $\texttt{friendsInfluence}_{\texttt{Preference}}(u, z)$, which relies on the `relevance` and `preference` functions (i.e., Equation 7 and Equation 8), but instead of the information of the user herself, the information of her friends is taken into account. More specifically, the $\texttt{friendsInfluence}_{\texttt{Relevance}}(u, z)$ measures how much topic z is related to the textual content of u's friends. Formally:

$$\texttt{friendsInfluence}_{\texttt{Relevance}}(u, z) = \frac{1}{|F_u|} \sum_{v \in F_u} \texttt{relevance}(v, z) \tag{11}$$

where F_u is the list of u's friends (or followees) and includes the users who are followed by user u (Wang et al., 2014) and $\texttt{relevance}(u, z)$ is measured based on Equation 7.

Further, $\texttt{friendsInfluence}_{\texttt{Preference}}(u, z)$ measures the similarity of topic z to the topics of interest to u's friends. Formally:

$$\texttt{friendsInfluence}_{\texttt{Preference}}(u, z) = \frac{1}{|F_u|} \sum_{v \in F_u} \texttt{preference}(v, z) \tag{12}$$

where $\texttt{preference}(u, z)$ is measured based on Equation 8.

4.2.5. Temporal Features

Given the features previously defined based on user, topic and interaction, we additionally include the time dimension to examine the impact of time on the future interest prediction problem. Acknowledging the dynamic nature of users' interests, temporal aspects have been utilized for user modeling in online social networks (Piao and Breslin, 2016b; Fani et al., 2020). To consider the dynamic nature of users' interests in user interest modeling approaches, numerous researchers have concentrated on the application of time decay functions across the historical content of the user (Ahmed et al., 2011; Orlandi et al., 2012). Time decay functions are employed to calculate the weight of each interest, taking into account its age. The underlying rationale behind interest decay functions is to assign greater importance to more recent interests, thereby emphasizing their significance.

On this basis, given users' historical activities are divided into T discrete time intervals $1 \leq t \leq T$, we generate the temporal version of metrics introduced so far in this section. To do so, for a given metric, we first calculate the metric in each time interval t and then apply a time decay function over the scores. In accordance with time decay functions, the weight of each metric in each time interval is calculated by taking into account its age, i.e., a higher weight is assigned to more recent time intervals compared to the old ones. Following this approach, we generate the temporal versions of the corresponding metrics of the following features: popularity, prestige, influence, relevance, and preference. In Table 1, we report the temporal features in each category of features (highlighting the gray-colored features). For example, to incorporate the temporality of a user's behavior in order to calculate her `influence`, we define user temporal `influence` that calculates the influence of the user in each time interval t and then aggregates them by using an exponential decay function $\partial(.)$ to weigh recent activities of the user more. Formally, we define user temporal `influence` as follows:

$$\texttt{temporalInfluence}(u) = \partial(\{influence^t(u) \mid 1 \leq t \leq T\}) \tag{13}$$

where $influence^t(u)$ calculates the `influence` of user u in time interval t by considering only the posts that the user u has published or retweeted in time interval t, i.e. $\mathbb{M}_u^t$.

It is noted that we have not included temporal versions of some features such as `personality traits` or `susceptibility` in interest prediction as they exhibit consistent characteristics over time or are not influenced by temporal changes. For some topical features including `freshness` and `coherence` and `exclusivity`, temporality is not applicable.

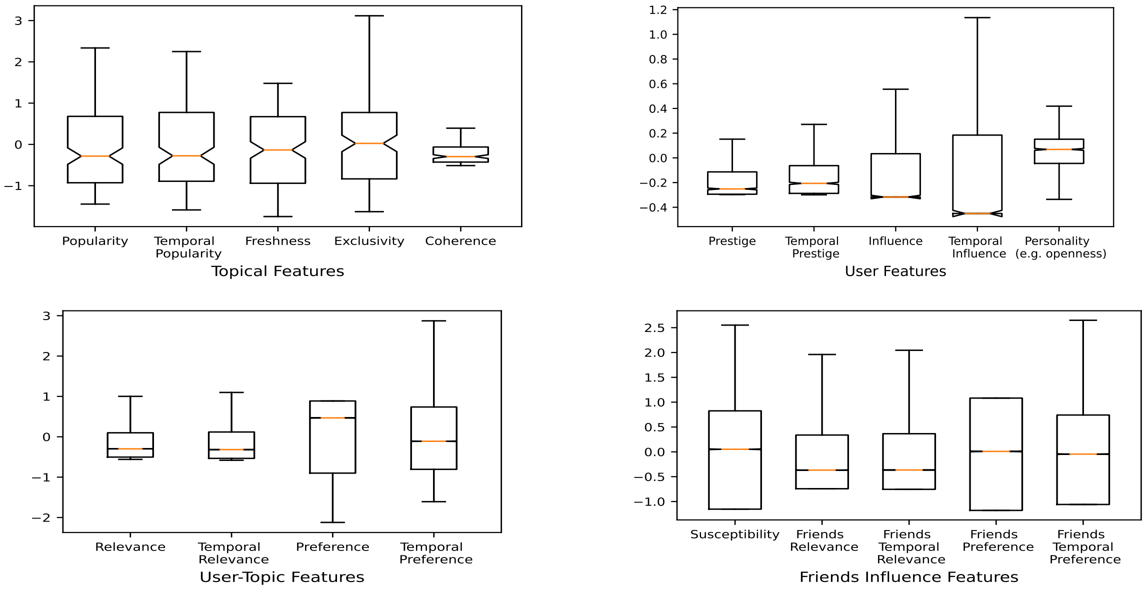

Figure 2: Feature distributions across different categories.

5. Empirical Evaluation

5.1. Dataset and Experimental Setup

Dataset. For experiments, a Twitter dataset used in (Zarrinkalam et al., 2019) is adopted and only considered those users who generated more than 100 tweets within 2-month period as golden users to ensure there is enough content per user. As a result of this process, a total of 1,556 distinct users were identified, and together they posted 1,607,585 tweets. The dataset also includes approximately 77,000 news articles that were obtained by crawling the URLs mentioned in the tweets. In accordance with the dataset under consideration, we extract the features utilizing the formula detailed in Section 4.2. The variable distribution of variable features is illustrated in Figure 2.

Topic Modeling. We divided the dataset into $T + 1$ time intervals in our experiments. In each time interval t, to extract K topics $\mathbb{Z}^t$, and topic profile of each user u, i.e., $P^t(u)$, we applied the implementation of Twitter-LDA[1] Zhao et al. (2011) on the collection of vocabularies found in the users' tweets that were posted in t, $\mathbb{M}^t$.

Time Interval and Topic Granularity. To set T, we follow the guidance by (Zarrinkalam et al., 2019) on the same dataset which has examined the impact of time interval length on the accuracy of predicting future interests. In their study, evaluations were conducted for various time interval lengths, including 1 day, 1 week, 2 weeks and 1 month and during the observations, it was noted that as the time interval length increased from 1 day to 1 week, the performance of interest prediction methods improved. This improvement can be attributed to the fact that longer time intervals contain richer data, which in turn enhances user interest detection and topic discovery. They have further found that as the length of the time interval exceeds 1 week, the diminishing temporal influence results in a decline in the quality of the prediction results. Therefore, the length of each time interval was set to 1 week for splitting the dataset into 8-time intervals. Regarding setting the number of topics in each time interval, i.e., K, we also follow the results of experiments done in (Zarrinkalam et al., 2019), which demonstrate that the conclusions derived from the results remain consistent across various numbers of topics in the majority of cases. Furthermore, the optimal performance was observed when the number of topics was set to $K = 20$. Consequently, with a time interval length of 1 week and

[1] https://github.com/minghui/Twitter-LDA

a total of 20 topics identified per week, the dataset was divided into eight-time intervals, resulting in a total of 160 topics. Moreover, similar to (Fang et al., 2016), we represented a topic $z \in \mathbb{Z}^t$ using a top-10 term representation.

Train-Test Split. Given the outputs of Twitter-LDA over the 8-time intervals of the dataset, the first 7 obtained topic profiles of the user u, $P^1(u), \ldots, P^7(u)$, are considered as her historical interests for training and the last topic profile, i.e. $P^8(u)$, as the user's future interests for testing.

Evaluation Metrics. Given $P^8(u)$ as our *ground truth*, we have used Normalized Discounted Cumulative Gain (NDCG) and Mean Average Precision (MAP) as two well-known Information Retrieval metrics to evaluate the ranking quality of different variants of our future interest prediction model and consequently answer our research questions. Additionally, to enhance the illustration of the model's performance, we present the results of experiments in terms of NDCG@5 and Mean Reciprocal Rank (MRR), in the Appendix. We note that in order to establish statistically robust claims regarding the importance of each feature in the future interest prediction task, we conduct a Wilcoxon test (Navidi, 2006) between the model incorporating all the features and the alternative variants. In our study, we aim to assess the statistical differences in performance metrics such as NDCG and MAP. This test is widely accepted in the literature for its effectiveness in detecting significant differences (Huang and Lei, 2023; Sanz-Cruzado et al., 2020) and is suitable for our experiments as we are evaluating the results of two distinct methods applied to the same set of users. The p-value derived from the test indicates whether the measurements of the two methods are statistically different. In all tables, symbols * and ** show statistical significance on a Wilcoxon test with p-value less than 0.01 and 0.05, respectively.

Textual Content of Users' Posts. Most researches on user interest modeling have only focused on using the social posts that a user interacts with (e.g., by liking, sharing, publishing) for modeling user interests as her textual content. However, since many posts include links (URLs) to embed more information about their topic, the textual content associated with these URLs has also been considered as an informative information source in some studies for detecting user's intentions and interests (Penas et al., 2013; Piao and Breslin, 2016c). Based on this, in the features which are extracted based on user content, we considered the textual content associated with a user as the combination of two content streams: 1) the collection of posts that she has originally posted or retweeted (Posts) and 2) the content of articles that she has mentioned their URLs in her posts (URLs).

User Activity Levels. To zoom in on the effectiveness of the activity level of users on answering each research question, as suggested in (Zhu et al., 2013; Zheng et al., 2018; Chen and Pirolli, 2012), we grouped users based on their activeness level and investigated the effectiveness of each afore-mentioned feature in future interest prediction in light of the user's activity. We computed the activity level of a user as the ratio of all posts published by the user compared to the total number of posts on social networks, Formally, $activeness(u) = |\mathbb{M}_u|/|\mathbb{M}|$, where $\mathbb{M}_u$ denotes the posts published by user u and $\mathbb{M}$ is all other posts in the network (Budak et al., 2014). We ranked users based on their activity level and divided them into three equal groups 1) *low-active*, 2) *semi-active*; and 3) *highly-active* users.

5.2. Baselines

We formulated the future interest prediction problem as a learning-to-rank (LTR) task with the objective of ranking the potential topics of the future based on the historical interests of a user. To answer our research questions and empirically study the relevance and impact of different features, both individually and collectively, we opted to utilize feature-based LTR models. Feature-based LTR models can be categorized into pointwise (e.g., MART (Friedman, 2001), Random Forest (Breiman, 2001)), pairwise (e.g., RankNet (Burges et al., 2005), RankBoost (Freund et al., 2003)), or listwise (e.g., ListNet (Cao et al., 2007), LambdaMART (Wu et al., 2010)) approaches. The categorization is based on their distinct approaches in handling training data and loss functions, each exhibiting unique strategies for leveraging features to enhance the ranking process (Deveaud et al., 2018).

It is noteworthy that we excluded LTR models relying on the original query-independent loss function (e.g., RankNet (Burges et al., 2005)), as these models have limitations in assessing query-level features. The evaluation results of selected LTR models on our dataset, measured in terms of NDCG and MAP, are presented in Table 2. We observed

Table 2
NDCG and MAP on test data reported by different learning-to-rank models.

Model	NDCG	MAP
Coordinate Ascent (Metzler and Bruce Croft, 2007)	0.4819	0.6411
MART (Friedman, 2001)	0.4753	0.6316
LamdaMART (Wu et al., 2010)	**0.4936**	**0.6586**
Random Forests (Breiman, 2001)	0.4869	0.6347
AdaRank (Xu and Li, 2007)	0.4524	0.6342
ListNet (Cao et al., 2007)	0.4806	0.6307

that LambdaMART emerged as the top-performing method for our task, surpassing other models in both NDCG and MAP. This observation aligns with the trend in the current LTR literature, where state-of-the-art methods often lean towards gradient-boosted decision trees, with LambdaMART being a prominent algorithm in this category (Lyzhin et al., 2023). Given these insights, we chose LambdaMART to report our findings for the remaining experiments. For LTR methods, we adopt the implementation in the RankLib[2] library.

5.3. Findings

We systematically present our findings on the six research questions introduced earlier in Section 4. To answer the research questions and to explore the relative effectiveness of each feature in each category, an ablation study is executed where we remove each feature at a time and retrain the model. Tables 3 and 4 show the results of our feature analysis for future interest prediction grouped by feature categories across different groups of users. We further summarize the evaluation result on each feature category in Tables 5 and 6.

5.3.1. RQ1: The Impact of User-based Features

In RQ1, we aim to understand the impact of user-based features on future interest prediction. Tables 3 and 4 show the result of our ablation study, in which we highlight statistically significant and impactful features for different groups of users. We observed that `influence` features positively impact the future interest prediction task, consistently across all groups of users. This is explained by the notion of positive reinforcement (Yoon and Tourassi, 2014) where users in social media receive immediate rewards in the form of attention from others serving as a useful predictor of their future behavior (in the context of our study, choosing their future topic of interest). In particular, we notice that users with low levels of activity in social media are more influenced by the ratio of attention that they receive from other users to select their future topic of interest compared to the other groups (see Table 3, removing this feature decreases NDCG significantly by 6.32% for low-activity users and 3.34% and 2.31% for semi-active/highly-active users.).

Our findings also show the effect of the user `prestige` feature, which refers to the position of a user in a social network in predicting the future interest for semi-active and highly-active users. We observed an improvement in the prediction outcome for these two user groups for both MAP and NDCG metrics, which implies that the more a user actively connects to other (prestigious) users, the higher the impact of such a feature is on her future interest prediction. This finding aligns with the earlier study (Yang et al., 2012), indicating that user `prestige` retains significant predictive strength in predicting the future usage of hashtags that a user has not previously employed.

Finally, we observed the importance of `personality traits` in modeling user's future interests for more active users. Specifically, Table 4 shows the significant improvement of 5.34% and 3.04% in MAP metric for semi-active and highly-active users. This finding confirms previous findings in the literature where `personality traits` are shown

[2]https://sourceforge.net/p/lemur/wiki/RankLib/

Table 3
Ablation study results for showing the importance of each feature for three groups of users in terms of the NDCG metric.

		Low-active users	Δ	Semi-active users	Δ	Highly-active users	Δ
	All features	**0.4797**		**0.4938**		**0.5072**	
Topical features	**- Popularity**	0.4623**	-3.62 %	0.4872**	-1.33 %	0.4821**	-4.95 %
	- Freshness	0.4513**	-5.92 %	0.4873*	-1.3 %	0.5023	-0.98 %
	- Exclusivity	0.4783	-0.29 %	0.4948	+0.21 %	0.5046	-0.53 %
	- Coherence	0.4616**	-3.76 %	0.4821**	-2.37 %	0.504	-0.64 %
User-based features	**- Prestige**	0.4898	+2.11 %	0.4796**	-2.87 %	0.4822**	-4.94 %
	- Influence	0.4494**	-6.32 %	0.4773**	-3.34 %	0.4955**	-2.31 %
	- Personality	0.4855	+1.22 %	0.4858**	-1.62 %	0.5004**	-1.36 %
Explicit user-topic engagement features	**- Relevance**	0.4379**	-8.72 %	0.4521**	-8.45 %	0.4639**	-8.55 %
	- Preference	0.4836	+0.81 %	0.4978	+0.81 %	0.5035	-0.74 %
Friends' influence features	**- Susceptibility**	0.4925	+2.68 %	0.5033	+1.93 %	0.5062	-0.21 %
	- Friends relevance	0.4922	+2.61 %	0.4827**	-2.24 %	0.4835**	-4.67 %
	- Friends preference	0.4859	+1.29 %	0.5005	+1.36 %	0.5029	-0.87 %

Table 4
Ablation study results for showing the importance of each feature for three groups of users in terms of MAP.

		Low-active users	Δ	Semi-active users	Δ	Highly-active users	Δ
	All features	**0.4481**		**0.6825**		**0.8452**	
Topical features	**- Popularity**	0.4189**	-6.53 %	0.663**	-2.86 %	0.8212**	-2.84 %
	- Freshness	0.4158**	-7.21 %	0.6609**	-3.16 %	0.8353**	-1.17 %
	- Exclusivity	0.4646**	+3.68 %	0.6743*	-1.2 %	0.8394	-0.69 %
	- Coherence	0.4322**	-3.55 %	0.6562**	-3.85 %	0.8331**	-1.43 %
User-based features	**- Prestige**	0.4702**	+4.91 %	0.6462**	-5.32 %	0.812**	-3.93 %
	- Influence	0.4125**	-7.95 %	0.629**	-7.84 %	0.8167**	-3.37 %
	- Personality	0.461	+2.87 %	0.6461**	-5.34 %	0.8193**	-3.07 %
Explicit user-topic engagement features	**- Relevance**	0.3975**	-11.3 %	0.6102**	-10.59 %	0.7913**	-6.37 %
	- Preference	0.4776**	+6.58 %	0.687	+0.66 %	0.8478	+0.31 %
Friends' influence features	**- Susceptibility**	0.4706**	+5.0 %	0.6783*	-0.62 %	0.8473*	+0.25 %
	- Friends relevance	0.4667**	+4.15 %	0.6279**	-8.0 %	0.7956**	-5.87 %
	- Friends preference	0.4725**	+5.44 %	0.6933**	+1.59 %	0.8511**	+0.7 %

to have predictive power in user interest modeling (Dhelim et al., 2020; Alrehili et al., 2022). The main findings from RQ1 can be summarized as follows:

First finding. Within user-specific characteristics, the positive effect of the `influence` feature consistently contributes to the task of predicting future interests among all user groups.

Table 5
Ablation study results for showing the importance of each category of features for three groups of users in terms of NDCG.

	Low-active users	Δ	Semi-active users	Δ	Highly-active users	Δ
All features	**0.4797**		**0.4938**		**0.5072**	
- Topical features	0.441**	-8.06 %	0.5024	+1.74 %	0.5014	-1.15 %
- User-based features	0.4728*	-1.43 %	0.4962	+0.49 %	0.4959**	-2.24 %
- Explicit user-topic engagement features	0.4488**	-6.44 %	0.4643**	-5.98 %	0.4628**	-8.75 %
- Friends' influence features	0.5138**	+7.11 %	0.5271**	+6.74 %	0.5052	-0.39 %
- Temporal features	0.4446**	-7.32 %	0.4689**	-5.03 %	0.5006*	-1.32 %

Table 6
Ablation study results for showing the importance of each category of features for three groups of users in terms of MAP.

	Low-active users	Δ	Semi-active users	Δ	Highly-active users	Δ
All features	**0.4481**		**0.6825**		**0.8452**	
- Topical features	0.4154**	-7.3 %	0.695**	+1.83 %	0.8613**	+1.91 %
- User-based features	0.4485	+0.08 %	0.6779*	-0.67 %	0.8422	-0.36 %
- Explicit user-topic engagement features	0.4232**	-5.57 %	0.621**	-9.0 %	0.8111**	-4.03 %
- Friends' influence features	0.4906**	+9.47 %	0.6842	+0.26 %	0.8201**	-2.97 %
- Temporal features	0.4015**	-10.4 %	0.6187**	-9.35 %	0.8114**	-4.0 %

Second finding. The effect of the user `prestige` feature on predicting future interests demonstrates its significance for both semi-active and highly-active users.

Third finding. When considering users who display greater levels of activity, it becomes evident that incorporating `personality traits` into the modeling process holds significant importance for accurately predicting their future interests.

5.3.2. RQ2: The Impact of Topical Features

To answer RQ2, i.e., what characteristics of topics affect predicting the future interest of users, from Tables 3 and 4, we notice that topic `popularity` that measures how much a topic is being discussed across the community is one of the strongest indicators for a user's future interest across different user groups. Interestingly, the impact is shown to be maximum for users with low activeness levels in social networks (compare the MAP reduction of 6.53% for low-activity users vs 2.86% and 2.84% for semi-active/highly-active users, after removing the feature). This can be explained by the viral nature of information dissemination on social networks, which makes even inactive silent users become interested in trendy topics (Mathioudakis and Koudas, 2010; Edelmann, 2016).

Besides topic `popularity`, we also observe that features such as topic `freshness` and topic `coherence` are more impactful for low-activity users in predicting their future interest across both MAP and NDCG metrics (see Tables 3 and 4). This shows that low-active users are more susceptible to being attracted to the newly emergent topic than the more active users. Our results also showed that the quality of the topic and its meaningfulness (aka topic `coherence`) is also impactful in attracting their attention in the future. This finding shares similarities with the research on topic coherence-based cross-domain recommender systems (Saraswat and Chakraverty, 2022). In their study, they utilize topic modeling on user-generated data and subsequently apply topic `coherence` computation, leading to an observed increase in precision within the recommender system.

Overall, our findings suggest that user-agnostic features in the topic category are better determinants in predicting the future interest of low-activity users (see Tables 5 and 6), which address RQ2. Yet, we were unable to find a strong correlation of topic `exclusivity` to the future interest prediction across different groups of users.

Fourth finding. For users with lower levels of activity, topical features such as topic `popularity`, `freshness`, and `coherence` emerge as powerful indicators for predicting their future interests.

5.3.3. RQ3: The Impact of User's Historical Interests

To address RQ3 and examine the impact of the user's historical interests on her future interests, we evaluate the importance of features that explicitly capture the relationship between users and topics in predicting users' future interests. We notice that `relevance`, which measures the relatedness of users' historical interest to the potential future topic relying on the frequency of terms mentioned in posts of the user, is the strongest predictor of users' future interest amongst features in all feature categories, and consistently across all groups of users based on both MAP and NDCG metrics (compare the NDCG decrease of 8.72% for low-activity users vs 8.45% and 8.55% for semi-active/highly-active users in Table 3). This highlights how the historical interests of users can effectively determine the future interest of users even for low-activity users.

Another interesting observation is the impact of the `preference` feature in future interest prediction for different groups of users. Although this feature captures the semantic similarity of users' historical topical interests to future topics, it is shown to be an insignificant feature in predicting users' future topical interests. This matter is more severe for low-activity users as removing this feature significantly improves their MAP value by 6.58% (see Table 4). One possible explanation for this observation can be due to the nature of users' generated content on social networks that are often short and noisy which may aggravate the impact of this feature specifically in the condition where there is not sufficient textual content available.

Overall, similar to various studies conducted in diverse applications, such as retweet prediction (Zarrinkalam et al., 2018), personalized news recommendation (Abel et al., 2011a), user interest detection (Bhattacharya et al., 2014; Zarrinkalam et al., 2015; Trikha et al., 2018), we can conclude the user-topic feature category is the strongest indicator of future user interest within all user groups (see Tables 5 and 6).

Fifth finding. Among all user groups, the user-topic feature category assumes a pivotal role in predicting future user interests. Notably, the `relevance` feature within the explicit user-topic engagement category emerges as the strongest indicator for the user's future interest across all user groups.

5.3.4. RQ4: The Impact of Social Influence

In RQ4, we account for the relationship between users on social networks to understand the effect of social influence in shaping the interest of users. Our findings show that `friend's relevance` feature, which simply measures the relatedness of a user's future topic to her friends' generated content plays an important role in articulating future interest of highly-active users (both NDCG and MAP metrics significantly increased by 4.67% and 5.87% as depicted in Tables 3 and 4). This aligns well with social influence theory and reiterates previous studies on the impact of social relationships in shaping the behavior and interest of users on social networks (Chen et al., 2013; Bao et al., 2013; Spasojevic et al., 2014; Wang et al., 2014). However, we have seen an adverse effect of such a feature for low-activity users which can be explained in light of their limited number of posts and limited number of friends (removing this feature improves NDCG and MAP metrics significantly by 2.61% and 4.15% as depicted in Tables 3 and 4).

Although expected, we did not see any significant improvement in the impact of `friends' preference` feature across all groups. This observation is aligned with our findings in RQ3 where the effect of such a feature is reported to be insignificant across all groups of users.

Furthermore, our ablation study on the `susceptibility` feature verifies the negative and insignificant impact of this feature specifically for low-activity and highly-active users. This matter is worse for low-activity users as it is shown to highly degrade the prediction outcome if this feature is included (decrease MAP by 5% depicted in Table 4). Thus, we cannot confirm the findings of previous researches on the impact of user's `susceptibility` in predicting user's interest (Budak et al., 2014).

Overall, our findings indicate an adverse effect of social influence-related features in revealing users' future interests specifically for users with high levels of activities. Our result cannot confirm the positive impact of such feature category for highly-active users across both evaluation metrics either (see Tables 5 and 6).

Sixth finding. The future interests of highly active users are influenced by their social connections, highlighting the crucial role that social relationships play in shaping their interests. Notably, among the friends' influence category, the `friends relevance` feature emerges as a stronger indicator compared to `friends preference` and `susceptibility` for future interest prediction of these users.

5.3.5. RQ5: Impact of Temporal Features

Through RQ5, we seek to understand the impact of temporal features in predicting the future interest of users. Our observations from Tables 7 and 8 indicate that considering temporality in calculating different categories of features significantly improves prediction performance across all groups of users. The results of the ablation study indicate that temporal features are more impactful for low-activity users such that removing them significantly lowers the prediction performance of this group by 7.32% in NDCG while this value is 1.32% (and statistically insignificant) for highly-active users (see Table 5). Similar patterns are observable in terms of the MAP metrics indicated in Table 6.

However, our feature analysis on the importance of each feature category indicates that considering temporality in features like `relevance` from explicit user-topic engagement features category significantly improves the prediction performance by 5.15% in NDCG for highly-active users, which is the highest amongst all feature categories in this user group. This confirms the findings in previous user interest modeling studies (Ahmed et al., 2011; Orlandi et al., 2012; Piao and Breslin, 2016b) and signifies the fact that users' topical interest will change over time so ignoring this would adversely affect prediction performance.

As another interesting observation, contrary to our finding regarding the positive effect of topical features (i.e., `popularity`, `freshness`, `coherence`) in prediction accuracy (i.e., RQ1), we did not see any significant improvement on the impact of `temporal popularity` feature across all groups (see Tables 7 and 8). Our findings imply that users (especially less active ones) rely more on topical features such as `popularity`, `coherence`, and `freshness` in determining their future interests, rather than relying on the dynamic popularity of topics over time.

Overall, the significance of temporal features in predicting future interests among diverse user groups with varying levels of activity is clearly demonstrated, highlighting their positive impact in this regard (Table 5 and 6). Specifically, temporal features within the user-based, explicit user-topic engagement and friends' influence feature categories demonstrate a favorable effect on future topic prediction. Nevertheless, our findings indicate that temporal features within the topical category do not have a significant positive effect on accurately predicting users' future interests across all user groups (Table 7 and 8).

Table 7
Ablation study results (NDCG) for temporal features on three groups of users.

	Low-active users	Δ	Semi-active users	Δ	Highly-active users	Δ
All features	**0.4797**		**0.4938**		**0.5072**	
- temporal Topical features	0.4806	+0.19 %	0.4917	-0.43 %	0.499*	-1.62 %
- temporal User-based features	0.4559**	-4.95 %	0.4764**	-3.52 %	0.4892**	-3.57 %
- temporal Explicit user-topic engagement features	0.4722	-1.57 %	0.4816**	-2.48 %	0.4811**	-5.15 %
- temporal Friends' influence features	0.4536**	-5.43 %	0.4658**	-5.68 %	0.497*	-2.02 %

Table 8
Ablation study results (MAP) for temporal features on three groups of users.

	Low-active users	Δ	Semi-active users	Δ	Highly-active users	Δ
All features	**0.4481**		**0.6825**		**0.8452**	
- temporal Topical features	0.4549**	+1.52 %	0.6953**	+1.88 %	0.8543**	+1.07 %
- temporal User-based features	0.4236**	-5.48 %	0.6412**	-6.04 %	0.8108**	-4.07 %
- temporal Explicit user-topic engagement features	0.4457	-0.56 %	0.6435**	-5.71 %	0.8142**	-3.67 %
- temporal Friends' influence features	0.4139**	-7.64 %	0.6239**	-8.58 %	0.8204**	-2.94 %

Seventh finding. The positive impact of temporal features is evident across user groups with different levels of activity, underscoring their significance in future interest prediction of users. However, it is important to highlight that the *temporal topical* features do not exhibit a significant positive effect.

5.3.6. RQ6: Impactful Features for Future Interest Prediction on Cold Topics

By answering our first five research questions, we showed the relative effectiveness of different features and their categories in predicting future users' interests on all future topics for different types of users. However, considering the rapid and dynamic nature of topics in response to events, future interest prediction faces *cold* topics, i.e., new topics that have not been previously observed in the past. In RQ6, we seek to understand the relative effectiveness of different features in predicting users' interests over a set of cold topics that are yet to emerge. To do so, similar to (Zarrinkalam et al., 2019), for each topic z_i within the testing time interval, i.e. $z_i \in \mathbb{Z}^{T+1}$, we calculate the level of activity by assessing the occurrence of topics within the intervals of training time, i.e. $\mathbb{Z}^1, ..., \mathbb{Z}^T$. The calculation is carried out as follows:

$$activity(z_i) = \frac{1}{T}\sum_{t=1}^{T} MAX_{z_i \in \mathbb{Z}^t} S(z_i, z_j) \tag{14}$$

where $S(z_i, z_j)$ represents the similarity between two topics z_i and z_j, which is calculated by computing the cosine similarity between the entity weight distribution vectors of the respective entities. Afterward, the topics are sorted in ascending order based on the level of activity and we then label the initial 25% of those as cold topics.

From Tables 11 and 12, we observe that user-based features along with features that capture social connections of users are the top 2 feature categories with the highest impact in predicting user interest on cold topics. In particular,

Table 9
Ablation study results (NDCG) for cold topics.

		All topics	Δ	Cold topics	Δ
	All features	**0.4936**		**0.6292**	
Topical features	**- Popularity**	0.4772**	-3.31 %	0.6231*	-0.98 %
	- Freshness	0.4803**	-2.69 %	0.5995**	-4.73 %
	- Exclusivity	0.4926	-0.2 %	0.5859**	-6.89 %
	- Coherence	0.4826**	-2.23 %	0.6094**	-3.15 %
User-based features	**- Prestige**	0.4839**	-1.97 %	0.5898**	-6.26 %
	- Influence	0.4741**	-3.95 %	0.5723**	-9.05 %
	- Personality	0.4906**	-0.61 %	0.5707**	-9.3 %
Explicit user-topic engagement features	**- Relevance**	0.4513**	-8.57 %	0.6613**	+5.09 %
	- Preference	0.495	+0.28 %	0.5942**	-5.57 %
Friends' influence features	**- Susceptibility**	0.5007	+1.44 %	0.5701**	-9.4 %
	- Friends relevance	0.4862**	-1.5 %	0.599**	-4.81 %
	- Friends preference	0.4964	+0.57 %	0.5819**	-7.53 %

Table 10
Ablation study results (MAP) for cold topics.

		All topics	Δ	Cold topics	Δ
	All features	**0.6586**		**0.6912**	
Topical features	**- Popularity**	0.6343**	-3.68 %	0.6872	-0.58 %
	- Freshness	0.6373**	-3.23 %	0.656**	-5.1 %
	- Exclusivity	0.6594	+0.13 %	0.6386**	-7.61 %
	- Coherence	0.6405**	-2.75 %	0.6656**	-3.7 %
User-based features	**- Prestige**	0.6428**	-2.4 %	0.6389**	-7.57 %
	- Influence	0.6194**	-5.95 %	0.6158**	-10.91 %
	- Personality	0.6421**	-2.5 %	0.6147**	-11.06 %
Explicit user-topic engagement features	**- Relevance**	0.5997**	-8.95 %	0.7396**	+7.0 %
	- Preference	0.6708**	+1.85 %	0.6407**	-7.31 %
Friends' influence features	**- Susceptibility**	0.6654**	+1.03 %	0.6081**	-12.02 %
	- Friends relevance	0.6301**	-4.33 %	0.6513**	-5.77 %
	- Friends preference	0.6723**	+2.08 %	0.6265**	-9.35 %

we learned that `influence`, `personality`, `susceptibility`, `friend's preferences` and `exclusivity` are the top 5 strongest indicators for predicting users' interest on *cold* topics (see Tables 9 and 10). Interestingly, `susceptibility`, which has been shown to have a negative impact on users' interest prediction on all topics, is the most impactful feature (improves MAP and NDCG by 11.09% and 9.3%) in determining user interest on cold topics. On the contrary, we notice that `relevance`, which has been shown to be the strongest indicator for predicting user interests on all topics, has the worst impact on prediction performance over cold topics. As indicated in Section 4.2.3, the `relevance` feature depends on the frequency of terms seen in previous topics, which potentially explains the reason behind its negative effectiveness in predicting cold topics.

Table 11
Ablation study results for showing the importance of each feature category for cold topics in terms of NDCG.

	All topics	Δ	Cold topics	Δ
All features	**0.4936**		**0.6292**	
- Topical features	0.4816**	-2.43 %	0.6432**	+2.22 %
- User-based features	0.4883**	-1.07 %	0.5947**	-5.49 %
- Explicit user-topic engagement features	0.4586**	-7.08 %	0.6381**	+1.41 %
- Friends' influence features	0.5154**	+4.41 %	0.6165**	-2.02 %
- Temporal features	0.4714**	-4.5 %	0.6028**	-4.2 %

Table 12
Ablation study results for showing the importance of each feature category for cold topics in terms of MAP.

	All topics	Δ	Cold topics	Δ
All features	**0.6586**		**0.6912**	
- Topical features	0.6572	-0.21 %	0.7117**	+2.96 %
- User-based features	0.6562*	-0.37 %	0.6516**	-5.72 %
- Explicit user-topic engagement features	0.6185**	-6.09 %	0.7001	+1.29 %
- Friends' influence features	0.665	+0.97 %	0.6663**	-3.6 %
- Temporal features	0.6105**	-7.3 %	0.6529**	-5.54 %

One of the most important observations is the superior impact of `personality traits` in determining the future interest of users over emerging topics. Our results confirm the findings of recent studies on the predictive power of this feature in user interest modeling (Alrehili et al., 2022). Yet, our findings further advocate for the strong impact of personality traits in predicting user interest on unobserved topics. Another interesting observation is the high impact of social relationship of users (aka `friend's preference`) in user interest prediction over cold topics (improves MAP and NDCG by 9.35% and 7.53%). Our experimental results suggest that users are more likely to be influenced by their friends in becoming interested in cold topics compared with the case where the topics are already being discussed. Our findings also confirm the findings from previous literature in user interest modeling (Bischof and Airoldi, 2012) on the effectiveness of topic `exclusivity` on attracting the attention of users on social networks. Furthermore, we notice that `influence` which is linked to the attention users receive from their social connections is one of the strongest indicators for both experiments (i.e., all topics and cold topics). This again highlights the importance of positive reinforcement and users' position in their social networks in predicting their future interests.

Overall, from Tables 11 and 12, we can conclude that those feature categories that consider the topical information are impractical to predict users' interest over unobserved topics. As such, the explicit user-topic engagement features category in tandem with topical features are shown to have a negative impact on the prediction task. On the other hand, we learned that relying on user-based features and social connections and considering their dynamicity can significantly improve the prediction task for cold topics.

The main findings from RQ6 can be summarized as follows:

Eighth finding. The prediction task for *cold* topics can be significantly improved by incorporating two feature categories: user-based features, specifically `influence` and `personality traits`, and friends' influence features, with a particular focus on `susceptibility` and `friends preferences`.

Ninth finding. The `relevance` feature within the explicit user-topic engagement category has a detrimental impact on users' future interests over *cold* topics.

6. Discussion

The primary aim of this paper is to comprehensively investigate the prediction of future interests on social networks by evaluating the effectiveness of various feature types in four feature categories, namely user-based, topical, explicit user-topic engagement, and friends' influence. These categories are analyzed across three user groups, consisting of low-activity, semi-active, and highly-active users. Table 13 presents a comprehensive overview of the features that demonstrate noteworthy effectiveness in predicting interests for each user group across diverse feature categories. Furthermore, Table 14 identifies the specific feature category that has either a positive or negative impact for each user group. Based on the results, we observe that topical features, especially `popularity`, `freshness`, and `coherence`, exert a substantial influence on the interests of low-active and semi-active users. It is crucial to note that our findings clearly reveal the existence of popularity bias in social media (Yalcin and Bilge, 2022). The evidence indicates that topics with high current engagement levels tend to attract even more attention from users in the future, establishing a self-reinforcing feedback loop that amplifies the popularity of such content. Notably, our results underscore that popularity bias exerts a substantial influence on the interests of low-active users, while its impact on highly-active users is comparatively less pronounced. This differential effect showed the complexity of popularity bias and its varied implications across different user categories. Our intention is to ascertain the degree of bias present in these predictions, which we plan to address more comprehensively in our future research.

On the other hand, user category features, specifically `prestige`, `influence`, and `personality` demonstrate significant efficacy in predicting interests for highly-active users. Notably, the user-topic features, particularly `relevance`, exhibit effectiveness across all user groups. For *cold* topics, the analysis reveals that two specific categories, namely friends' influence and user-based features, have a positive impact on interest prediction. For feature categories that have a negative impact on interest prediction for specific user groups and topic types, such as cold topics, based on the results reported in Table 14, we observe that friends' influence features have a detrimental effect on the interests of low-activity and semi-active users, which is expected given the low degree of engagement from such users with the community. Furthermore, in the case of cold topics, we observe that both topical features and explicit user-topic engagement features show a negative impact. This implies that these feature categories are not effective in accurately predicting user interests for cold topics.

Table 15 offers a comprehensive summary of the temporal features within each feature category, showcasing their effectiveness in predicting interests for each user group. Taking the results into consideration, temporal features within the user-based, explicit user-topic engagement, and friends' influence categories have a beneficial effect on predicting future topics. It highlights the significance of considering the temporal dimensions in the interests of users when predicting their future interests. However, when examining the dynamicity of topic `popularity` over time, based on our results, we found it does not consistently improve the accuracy of predicting users' future interests across all user groups. It is probably because the `popularity` of topics in our 2-month dataset has not changed significantly. Therefore, the dynamic nature of users' interests plays a vital role in the accurate prediction of users' future interests, making it imperative to incorporate this dynamicity into future interest modeling approaches.

Table 13
Significant features for interest prediction in different user groups and feature categories.

	Low-active users	**Semi-active users**	**Highly-active users**
Topical features	Popularity Freshness Coherence	Popularity Freshness Coherence	Popularity
User-based features	Influence	Prestige Influence Personality	Prestige Influence Personality
Explicit user-topic engagement features	Relevance	Relevance	Relevance
Friends' influence features	-	Friends' relevance	Friends' relevance

Table 14
Positive and negative impact of feature categories for different user groups and cold topics, represented as ✓and ✘ symbols respectively.

	Low-active users	**Semi-active users**	**Highly-active users**	**Cold topics**
Topical features	✓	✘		✘
User-based features			✓	✓
Explicit user-topic engagement features	✓	✓	✓	✘
Friends' influence features	✘	✘	✓	✓

Table 15
Positive and negative impact of temporal feature categories for different user groups, represented as ✓and ✘ symbols respectively.

	Low-active users	**Semi-active users**	**Highly-active users**
temporal Topical features	✘		
temporal User-based features	✓	✓	✓
temporal Explicit user-topic engagement features		✓	✓
temporal Friends' influence features	✓	✓	✓

7. Conclusion and Future Work

In this paper, we have studied the problem of user interest prediction toward a set of unobserved topics in the future. The problem was formulated as a learning-to-rank task with the objective of ranking the potential topics of the future based on the historical user's interests. We defined 19 features that are hypothesized to have an impact on the quality of interest predictions and validated our hypothesis by analyzing the importance of features by doing an ablation study. After pairwise comparison between different groups of users with different activity levels in social networks, we observe significant differences in the importance of different features in different experimental settings. In summary, we found that `relevance` in the explicit user-topic engagement feature category is the most powerful indicator of a user's future interest, regardless of the user group. For low-activity users, topical features such as topic `popularity`, `freshness`, and `coherence` strongly predict their future interest. Moreover, temporal features demonstrate a positive impact on different user groups with varying levels of activity in social networks. Lastly, Incorporating user-based features, including `influence` and `personality`, along with friends' influence features, specifically `susceptibility` and `friends' preferences`, significantly improves the prediction accuracy for cold topics.

Based on this foundation, our future work will delve into several areas, one of which involves investigating the influence of these features on different user interest modeling strategies. As introduced in (Zarrinkalam et al., 2020), there exist three distinct types of user interest modeling: (1) explicit user interest detection; (2) implicit user interest inference and (3) future interest prediction. One potential venue for future work is to perform a comparative study on the effectiveness of these features in different user modeling strategies and understand the synergistic or conflicting effects of such features. Another area that we will explore is to investigate the impact of other internal and external sources of information used for user interest modeling on predicting future user interest. The features in the current study are only extracted from social posts and user relations, however, there are several studies on user interest modeling that have leveraged additional information such as list membership (Piao and Breslin, 2017) for this purpose. There are also some studies that incorporate Knowledge bases like Freebase and Wikipedia, to enrich social posts and improve the accuracy of user profiles. We intend to enrich our framework with additional features that are extracted based on these sources of information and investigate their effect on the future user interest prediction task. Finally, building on our findings that highlight the presence of popularity bias within social media and its impact on the future interests of low-active users, our intent is to explicitly quantify and address the extent of these biases. Inspired by Yalcin and Bilge (2022), our future work will involve a comprehensive exploration of methodologies and strategies aimed at mitigating these biases. This approach will be adopted not only for understanding but also for actively improving the fairness and accuracy of interest predictions across diverse users.

A. Appendix. Model Performance in Terms of NDCG@5 and MRR

In our experimental analysis, we calculate NDCG and MAP metrics by considering the entire ranked list. To specifically evaluate the model's performance at the upper segments of this list, we report NDCG@5 and MRR metrics in this section. Upon careful examination of the tables presented here and their corresponding representations in the paper, we observe the consistency in all nine findings between the metrics reported in this section and those presented in the paper.

Table 16
Ablation study results for showing the importance of each feature for three groups of users in terms of NDCG@5 metric.

		Low-active users	Δ	Semi-active users	Δ	Highly-active users	Δ
	All features	**0.2203**		**0.2275**		**0.261**	
Topical features	- Popularity	0.1798**	-18.37 %	0.2193*	-3.6 %	0.217**	-16.86 %
	- Freshness	0.1888**	-14.3 %	0.2236	-1.72 %	0.253	-3.06 %
	- Exclusivity	0.2216	+0.59 %	0.227	-0.24 %	0.2551	-2.27 %
	- Coherence	0.211	-4.22 %	0.2238	-1.64 %	0.2599	-0.41 %
User-based features	- Prestige	0.2324	+5.49 %	0.2095**	-7.92 %	0.224**	-14.79 %
	- Influence	0.175**	-20.58 %	0.2002**	-12.01 %	0.2371**	-9.16 %
	- Personality	0.2436**	+10.56 %	0.2262	-0.59 %	0.2485**	-4.78 %
Explicit user-topic engagement features	- Relevance	0.1641**	-25.5 %	0.1661**	-27.0 %	0.1921**	-26.38 %
	- Preference	0.2239	+1.62 %	0.2346	+3.09 %	0.2593	-0.65 %
Friends' influence features	- Susceptibility	0.2378	+7.93 %	0.2411	+5.97 %	0.2624	+0.54 %
	- Friends relevance	0.2463*	+11.8 %	0.2176	-4.39 %	0.2243**	-14.07 %
	- Friends preference	0.2357	+6.97 %	0.2462	+8.21 %	0.2582	-1.07 %

Table 17
Ablation study results for showing the importance of each feature for three groups of users in terms of MRR.

		Low-active users	Δ	Semi-active users	Δ	Highly-active users	Δ
	All features	**0.5753**		**0.7872**		**0.9246**	
Topical features	- Popularity	0.4898**	-14.86 %	0.7651	-2.8 %	0.8898**	-3.76 %
	- Freshness	0.4922**	-14.45 %	0.769	-2.3 %	0.932	+0.8 %
	- Exclusivity	0.5633	-2.08 %	0.7738	-1.7 %	0.9029*	-2.34 %
	- Coherence	0.4797**	-16.61 %	0.6678**	-15.05 %	0.8689**	-6.02 %
User-based features	- Prestige	0.6229**	+8.27 %	0.77	-2.18 %	0.9002*	-2.63 %
	- Influence	0.499**	-13.27 %	0.6885**	-12.54 %	0.8773**	-5.11 %
	- Personality	0.5834	+1.41 %	0.7405**	-5.93 %	0.898**	-2.87 %
Explicit user-topic engagement features	- Relevance	0.3771**	-34.46 %	0.5733**	-2717 %	0.7998**	-13.49 %
	- Preference	0.5591	-2.82 %	0.7622	-3.17 %	0.8946**	-3.24 %
Friends' influence features	- Susceptibility	0.6819**	+18.53 %	0.844**	+7.22 %	0.9331	+0.93 %
	- Friends relevance	0.6093**	+5.92 %	0.7052**	-10.41 %	0.8315**	-10.06 %
	- Friends preference	0.5899	+2.53 %	0.8173**	+3.83 %	0.931	+0.69 %

Table 18
Ablation study results for showing the importance of each category of features for three groups of users in terms of NDCG@5.

	Low-active users	Δ	Semi-active users	Δ	Highly-active users	Δ
All features	**0.2203**		**0.2275**		**0.261**	
- Topical features	0.1661**	-24.61 %	0.2419	+6.31 %	0.2551	-2.27 %
- User-based features	0.209	-5.12 %	0.2422	+6.45 %	0.2449*	-6.18 %
- Explicit user-topic engagement features	0.1749**	-20.63 %	0.1806**	-20.65 %	0.1886**	-27.72 %
- Friends' influence features	0.2575**	+16.87 %	0.2759**	+21.25 %	0.2579	-1.2 %
- Temporal features	0.1681**	-23.7 %	0.1892**	-16.83 %	0.2499	-4.23 %

Table 19
Ablation study results for showing the importance of each category of features for three groups of users in terms of MRR.

	Low-active users	Δ	Semi-active users	Δ	Highly-active users	Δ
All features	**0.5753**		**0.7872**		**0.9246**	
- Topical features	0.5153**	-10.43 %	0.8195**	+4.11 %	0.9659**	+4.48 %
- User-based features	0.5985	+4.03 %	0.8147*	+3.5 %	0.9328	+0.89 %
- Explicit user-topic engagement features	0.4484**	-22.06 %	0.6383**	-18.91 %	0.8556**	-7.46 %
- Friends' influence features	0.7648**	+32.95 %	0.9465**	+20.24 %	0.972**	+5.13 %
- Temporal features	0.5112**	-11.13 %	0.7141**	-9.28 %	0.9063	-1.98 %

Table 20
Ablation study results (NDCG@5) for temporal features on three groups of users.

	Low-active users	Δ	Semi-active users	Δ	Highly-active users	Δ
All features	**0.2203**		**0.2275**		**0.261**	
- temporal Topical features	0.2229	+1.17 %	0.2327	+2.26 %	0.2466	-5.5 %
- temporal User-based features	0.1846**	-16.21 %	0.1991**	-12.49 %	0.2287**	-12.36 %
- temporal Explicit user-topic engagement features	0.2074	-5.86 %	0.2108	-7.34 %	0.2211**	-15.27 %
- temporal Friends' influence features	0.1752**	-20.47 %	0.1841**	-19.09 %	0.2451*	-6.08 %

Table 21
Ablation study results (MRR) for temporal features on three groups of users.

	Low-active users	Δ	Semi-active users	Δ	Highly-active users	Δ
All features	**0.5753**		**0.7872**		**0.9246**	
- temporal Topical features	0.548	-4.74 %	0.7803	-0.88 %	0.9096	-1.62 %
- temporal User-based features	0.4815**	-16.31 %	0.7199**	-8.55 %	0.8733**	-5.54 %
- temporal Explicit user-topic engagement features	0.5298**	-7.91 %	0.7464**	-5.18 %	0.8833**	-4.41 %
- temporal Friends' influence features	0.5371**	-6.65 %	0.702**	-10.82 %	0.9048*	-2.13 %

Table 22
Ablation study results (NDCG@5) for cold topics.

		All topics	Δ	Cold topics	Δ
	All features	**0.2363**		**0.6292**	
Topical features	**- Popularity**	0.2054**	-13.08 %	0.6231*	-0.98 %
	- Freshness	0.2218**	-6.12 %	0.5995**	-4.73 %
	- Exclusivity	0.2346	-0.73 %	0.5859**	-6.89 %
	- Coherence	0.2316*	-1.99 %	0.6094**	-3.15 %
User-based features	**- Prestige**	0.2214**	-6.28 %	0.5898**	-6.26 %
	- Influence	0.2041**	-13.63 %	0.5723**	-9.05 %
	- Personality	0.2394	+1.33 %	0.5707**	-9.3 %
Explicit user-topic engagement features	**- Relevance**	0.1741**	-26.31 %	0.6613**	+5.09 %
	- Preference	0.2392	+1.25 %	0.5942**	-5.57 %
Friends' influence features	**- Susceptibility**	0.2471	+4.58 %	0.5701**	-9.4 %
	- Friends relevance	0.2294*	-2.92 %	0.599**	-4.81 %
	- Friends preference	0.2467	+4.41 %	0.5819**	-7.53 %

Table 23
Ablation study results (MRR) for cold topics.

		All topics	Δ	Cold topics	Δ
	All features	**0.7623**		**0.7436**	
Topical features	**- Popularity**	0.7149**	-6.22 %	0.7362	-1.04 %
	- Freshness	0.731**	-4.1 %	0.7054**	-5.13 %
	- Exclusivity	0.7467*	-2.06 %	0.6665**	-10.36 %
	- Coherence	0.6725**	-11.79 %	0.7172**	-3.54 %
User-based features	**- Prestige**	0.7644	+0.27 %	0.6814**	-8.35 %
	- Influence	0.6883**	-9.72 %	0.6297**	-15.31 %
	- Personality	0.7407**	-2.84 %	0.6348**	-14.63 %
Explicit user-topic engagement features	**- Relevance**	0.5834**	-23.47 %	0.81**	+8.94 %
	- Preference	0.7386**	-3.11 %	0.6643**	-10.65 %
Friends' influence features	**- Susceptibility**	0.8196**	+7.52 %	0.624**	-16.08 %
	- Friends relevance	0.7154**	-6.16 %	0.6933**	-6.76 %
	- Friends preference	0.7794*	+2.23 %	0.6626**	-10.88 %

Table 24
Ablation study results for showing the importance of each feature category for cold topics in terms of NDCG@5.

	All topics	Δ	Cold topics	Δ
All features	**0.2363**		**0.6292**	
- Topical features	0.221**	-6.46 %	0.6432**	+2.22 %
- User-based features	0.232*	-1.8 %	0.5947**	-5.49 %
- Explicit user-topic engagement features	0.1814**	-23.25 %	0.6381	+1.41 %
- Friends' influence features	0.2637**	+11.62 %	0.6165**	-2.02 %
- Temporal features	0.2024**	-14.32 %	0.6028**	-4.2 %

Table 25
Ablation study results for showing the importance of each feature category for cold topics in terms of MRR.

	All topics	Δ	Cold topics	Δ
All features	**0.7623**		**0.7436**	
- Topical features	0.7669	+0.6 %	0.7698**	+3.53 %
- User-based features	0.7819**	+2.57 %	0.688**	-7.47 %
- Explicit user-topic engagement features	0.6475**	-15.07 %	0.7731**	+3.97 %
- Friends' influence features	0.8944**	+17.33 %	0.728**	-2.14 %
- Temporal features	0.7105**	-6.79 %	0.6932**	-6.77 %

References

Abel, F., Gao, Q., Houben, G.J., Tao, K., 2011a. Analyzing user modeling on twitter for personalized news recommendations, in: User Modeling, Adaption and Personalization: 19th International Conference, UMAP 2011, Girona, Spain, July 11-15, 2011. Proceedings 19, Springer. pp. 1–12.

Abel, F., Gao, Q., Houben, G.J., Tao, K., 2011b. Semantic enrichment of twitter posts for user profile construction on the social web, in: The Semanic Web: Research and Applications: 8th Extended Semantic Web Conference, ESWC 2011, Heraklion, Crete, Greece, May 29–June 2, 2011, Proceedings, Part II 8, Springer. pp. 375–389.

Ahmed, A., Low, Y., Aly, M., Josifovski, V., Smola, A.J., 2011. Scalable distributed inference of dynamic user interests for behavioral targeting, in: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 114–122.

Alavijeh, S.Z., Zarrinkalam, F., Noorian, Z., Mehrpour, A., Etminani, K., 2023. What users' musical preference on twitter reveals about psychological disorders. Information Processing & Management 60, 103269.

Aldous, K.K., An, J., Jansen, B.J., 2023. What really matters?: characterising and predicting user engagement of news postings using multiple platforms, sentiments and topics. Behaviour & Information Technology 42, 545–568.

Alrehili, M.M., Yafooz, W.M., Alsaeedi, A., Emara, A.H.M., Saad, A., Al Aqrabi, H., 2022. The impact of personality and demographic variables in collaborative filtering of user interest on social media. Applied Sciences 12, 2157.

Bao, H., Li, Q., Liao, S.S., Song, S., Gao, H., 2013. A new temporal and social pmf-based method to predict users' interests in micro-blogging. Decision Support Systems 55, 698–709.

Bennacer Seghouani, N., Jipmo, C.N., Quercini, G., 2019. Determining the interests of social media users: two approaches. Information Retrieval Journal 22, 129–158.

Bhattacharya, P., Zafar, M.B., Ganguly, N., Ghosh, S., Gummadi, K.P., 2014. Inferring user interests in the twitter social network, in: Proceedings of the 8th ACM Conference on Recommender systems, pp. 357–360.

Bischof, J., Airoldi, E.M., 2012. Summarizing topical content with word frequency and exclusivity, in: Proceedings of the 29th international conference on machine learning (icml-12), pp. 201–208.

Breiman, L., 2001. Random forests. Machine learning 45, 5–32.

Brin, S., 1998. The pagerank citation ranking: bringing order to the web. Proceedings of ASIS, 1998 98, 161–172.

Budak, C., Kannan, A., Agrawal, R., Pedersen, J., 2014. Inferring user interests from microblogs. AAAI ICWSM .

Burges, C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N., Hullender, G., 2005. Learning to rank using gradient descent, in: Proceedings of the 22nd international conference on Machine learning, pp. 89–96.

Cao, Z., Qin, T., Liu, T.Y., Tsai, M.F., Li, H., 2007. Learning to rank: from pairwise approach to listwise approach, in: Proceedings of the 24th international conference on Machine learning, pp. 129–136.

Cha, M., Haddadi, H., Benevenuto, F., Gummadi, K., 2010. Measuring user influence in twitter: The million follower fallacy, in: Proceedings of the international AAAI conference on web and social media, pp. 10–17.

Chen, J., Pirolli, P., 2012. Why you are more engaged: factors influencing twitter engagement in occupy wall street, in: Proceedings of the International AAAI Conference on Web and Social Media, pp. 423–426.

Chen, W., Hsu, W., Lee, M.L., 2013. Modeling user's receptiveness over time for recommendation, in: Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval, pp. 373–382.

Cheng, L., Shi, Y., Li, L., Yu, H., Wang, X., Yan, Z., 2023. Kleca: knowledge-level-evolution and category-aware personalized knowledge recommendation. Knowledge and Information Systems 65, 1045–1065.

Das, A., Roy, M., Dutta, S., Ghosh, S., Das, A.K., 2015. Predicting trends in the twitter social network: a machine learning approach, in: Swarm, Evolutionary, and Memetic Computing: 5th International Conference, SEMCCO 2014, Bhubaneswar, India, December 18-20, 2014, Revised Selected Papers 5, Springer. pp. 570–581.

Deng, L., Jia, Y., Zhou, B., Huang, J., Han, Y., 2018. User interest mining via tags and bidirectional interactions on sina weibo. World Wide Web 21, 515–536.

Deveaud, R., Mothe, J., Ullah, M.Z., Nie, J.Y., 2018. Learning to adaptively rank document retrieval system configurations. ACM Transactions on Information Systems (TOIS) 37, 1–41.

Dhelim, S., Aung, N., Ning, H., 2020. Mining user interest based on personality-aware hybrid filtering in social networks. Knowledge-Based Systems 206, 106227.

Di Tommaso, G., Faralli, S., Stilo, G., Velardi, P., 2018. Wiki-mid: a very large multi-domain interests dataset of twitter users with mappings to wikipedia, in: The Semantic Web–ISWC 2018: 17th International Semantic Web Conference, Monterey, CA, USA, October 8–12, 2018, Proceedings, Part II 17, Springer. pp. 36–52.

Edelmann, N., 2016. 13. what is lurking? a literature review of research on lurking. The psychology of social networking 1, 159–174.

Fang, A., Macdonald, C., Ounis, I., Habel, P., 2016. Topics in tweets: A user study of topic coherence metrics for twitter data, in: Advances in Information Retrieval: 38th European Conference on IR Research, ECIR 2016, Padua, Italy, March 20–23, 2016. Proceedings 38, Springer. pp. 492–504.

Fani, H., Bashari, M., Zarrinkalam, F., Bagheri, E., Al-Obeidat, F., 2018. Stopword detection for streaming content, in: Advances in Information Retrieval: 40th European Conference on IR Research, ECIR 2018, Grenoble, France, March 26-29, 2018, Proceedings 40, Springer. pp. 737–743.

Fani, H., Jiang, E., Bagheri, E., Al-Obeidat, F., Du, W., Kargar, M., 2020. User community detection via embedding of social network structure and temporal content. Information Processing & Management 57, 102056.

Freire, M., Antunes, F., Costa, J.P., 2022. Getting decision support from context-specific online social networks: a case study. Social Network Analysis and Mining 12, 41.

Freund, Y., Iyer, R., Schapire, R.E., Singer, Y., 2003. An efficient boosting algorithm for combining preferences. Journal of machine learning research 4, 933–969.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Annals of statistics , 1189–1232.

Galal, S., Nagy, N., El-Sharkawi, M.E., 2021. Cnmf: A community-based fake news mitigation framework. Information 12, 376.

Gayo-Avello, D., 2013. Nepotistic relationships in twitter and their impact on rank prestige algorithms. Information Processing & Management 49, 1250–1280.

Hassan, N., El-Sharkawi, M.E., El-Tazi, N., 2016. Measuring user's susceptibility to influence in twitter, in: Social Data Analytics and Management Workshop, co-located with VLDB.

He, W., Liu, H., He, J., Tang, S., Du, X., 2015. Extracting interest tags for non-famous users in social network, in: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, pp. 861–870.

Huang, D., Lei, F., 2023. Temporal group-aware graph diffusion networks for dynamic link prediction. Inf. Process. Manag. 60, 103292. URL: https://doi.org/10.1016/j.ipm.2023.103292, doi:10.1016/J.IPM.2023.103292.

Inaba, M., Takahashi, K., 2018. Estimating user interest from open-domain dialogue, in: Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue, pp. 32–40.

Jamali, M., Ester, M., 2010. A matrix factorization technique with trust propagation for recommendation in social networks, in: Proceedings of the fourth ACM conference on Recommender systems, pp. 135–142.

Jipmo, C.N., Quercini, G., Bennacer, N., 2017. Frisk: a multilingual approach to find twitter interests via wikipedia, in: Advanced Data Mining and Applications: 13th International Conference, ADMA 2017, Singapore, November 5–6, 2017, Proceedings 13, Springer. pp. 243–256.

Kacem, A., Boughanem, M., Faiz, R., 2014. Time-sensitive user profile for optimizing search personlization, in: User Modeling, Adaptation, and Personalization: 22nd International Conference, UMAP 2014, Aalborg, Denmark, July 7-11, 2014. Proceedings 22, Springer. pp. 111–121.

Kanavos, A., Livieris, I.E., 2020. Fuzzy information diffusion in twitter by considering user's influence. International Journal on Artificial Intelligence Tools 29, 2040003.

Kang, J., Choi, H., Lee, H., 2019. Deep recurrent convolutional networks for inferring user interests from social media. Journal of Intelligent Information Systems 52, 191–209.

Kapanipathi, P., Jain, P., Venkataramani, C., Sheth, A., 2014. User interests identification on twitter using a hierarchical knowledge base, in: The Semantic Web: Trends and Challenges: 11th International Conference, ESWC 2014, Anissaras, Crete, Greece, May 25-29, 2014. Proceedings 11, Springer. pp. 99–113.

Kapanipathi, P., Orlandi, F., Sheth, A.P., Passant, A., 2011. Personalized filtering of the twitter stream .

Karatay, D., Karagoz, P., 2015. User interest modeling in twitter with named entity recognition, in: 5th Workshop on Making Sense of Microposts.

Kursuncu, U., Gaur, M., Lokala, U., Thirunarayan, K., Sheth, A., Arpinar, I.B., 2019. Predictive analysis on twitter: Techniques and applications. Emerging research challenges and opportunities in computational social network analysis and mining , 67–104.

Lampropoulos, G., Anastasiadis, T., Siakas, K., Siakas, E., 2022. The impact of personality traits on social media use and engagement: An overview. International Journal on Social and Education Sciences 4, 34–51.

Lee, R.K.W., Lim, E.P., 2015. Measuring user influence, susceptibility and cynicalness in sentiment diffusion, in: European Conference on Information Retrieval, Springer. pp. 411–422.

Liu, J., Xiao, Y., Zheng, W., Hsu, C.H., 2023. Siga: social influence modeling integrating graph autoencoder for rating prediction. Applied Intelligence 53, 6432–6447.

Lyzhin, I., Ustimenko, A., Gulin, A., Prokhorenkova, L., 2023. Which tricks are important for learning to rank?, in: International Conference on Machine Learning, PMLR. pp. 23264–23278.

Mathioudakis, M., Koudas, N., 2010. Twittermonitor: trend detection over the twitter stream, in: Proceedings of the 2010 ACM SIGMOD International Conference on Management of data, pp. 1155–1158.

McCrae, R.R., John, O.P., 1992. An introduction to the five-factor model and its applications. Journal of personality 60, 175–215.

McPherson, M., Smith-Lovin, L., Cook, J.M., 2001. Birds of a feather: Homophily in social networks. Annual review of sociology 27, 415–444.

Metzler, D., Bruce Croft, W., 2007. Linear feature-based models for information retrieval. Information Retrieval 10, 257–274.

Mislove, A., Viswanath, B., Gummadi, K.P., Druschel, P., 2010. You are who you know: inferring user profiles in online social networks, in: Proceedings of the third ACM international conference on Web search and data mining, pp. 251–260.

Musiał, K., Kazienko, P., Brodka, P., 2009. User position measures in social networks, in: Proceedings of the 3rd workshop on social network mining and analysis, pp. 1–9.

Navidi, W.C., 2006. Statistics for engineers and scientists. volume 2. McGraw-Hill New York.

Okuda, Y., Sudoh, K., Shinagawa, S., Nakamura, S., 2023. Modeling multiple user interests using hierarchical knowledge for conversational recommender system. arXiv preprint arXiv:2303.00311 .

On-At, S., Quirin, A., Péninou, A., Baptiste-Jessel, N., Canut, M.F., Sèdes, F., 2017. A parametric study to construct time-aware social profiles. Trends in Social Network Analysis: Information Propagation, User Behavior Modeling, Forecasting, and Vulnerability Assessment , 21–50.

Orlandi, F., Breslin, J., Passant, A., 2012. Aggregated, interoperable and multi-domain user profiles for the social web, in: Proceedings of the 8th International Conference on Semantic Systems, pp. 41–48.

Penas, P., Del Hoyo, R., Vea-Murguía, J., González, C., Mayo, S., 2013. Collective knowledge ontology user profiling for twitter–automatic user profiling, in: 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), IEEE. pp. 439–444.

Pennacchiotti, M., Silvestri, F., Vahabi, H., Venturini, R., 2012. Making your interests follow you on twitter, in: Proceedings of the 21st ACM international conference on Information and knowledge management, pp. 165–174.

Pereira, F.S., Gama, J., de Amo, S., Oliveira, G.M., 2018. On analyzing user preference dynamics with temporal social networks. Machine Learning 107, 1745–1773.

Perifanis, V., Drosatos, G., Stamatelatos, G., Efraimidis, P.S., 2023. Fedpoirec: Privacy-preserving federated poi recommendation with social influence. Information Sciences 623, 767–790.

Piao, G., Breslin, J.G., 2016a. Analyzing aggregated semantics-enabled user modeling on google+ and twitter for personalized link recommendations, in: proceedings of the 2016 conference on user modeling adaptation and personalization, pp. 105–109.

Piao, G., Breslin, J.G., 2016b. Exploring dynamics and semantics of user interests for user modeling on twitter for link recommendations, in: proceedings of the 12th international conference on semantic systems, pp. 81–88.

Piao, G., Breslin, J.G., 2016c. Interest representation, enrichment, dynamics, and propagation: a study of the synergetic effect of different user modeling dimensions for personalized recommendations on twitter, in: European Knowledge Acquisition Workshop, Springer. pp. 496–510.

Piao, G., Breslin, J.G., 2017. Leveraging followee list memberships for inferring user interests for passive users on twitter, in: Proceedings of the 28th ACM Conference on Hypertext and Social Media, pp. 155–164.

Qi, T., Wu, F., Wu, C., Huang, Y., 2022a. Fum: fine-grained and fast user modeling for news recommendation, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1974–1978.

Qi, T., Wu, F., Wu, C., Huang, Y., 2022b. News recommendation with candidate-aware user modeling, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1917–1921.

Qi, T., Wu, F., Wu, C., Yang, P., Yu, Y., Xie, X., Huang, Y., 2021. Hierec: Hierarchical user interest modeling for personalized news recommendation. arXiv preprint arXiv:2106.04408 .

Qiu, Z., Hu, Y., Wu, X., 2022. Graph neural news recommendation with user existing and potential interest modeling. ACM Transactions on Knowledge Discovery from Data (TKDD) 16, 1–17.

Ribeiro, M., Calais, P., Santos, Y., Almeida, V., Meira Jr, W., 2018. Characterizing and detecting hateful users on twitter, in: Proceedings of the International AAAI Conference on Web and Social Media.

Riquelme, F., González-Cantergiani, P., 2016. Measuring user influence on twitter: A survey. Information processing & management 52, 949–975.

Rong, X., Mei, Q., 2013. Diffusion of innovations revisited: from social network to innovation network, in: Proceedings of the 22nd ACM international conference on Information & Knowledge Management, pp. 499–508.

Salminen, J., Mustak, M., Corporan, J., Jung, S.g., Jansen, B.J., 2022. Detecting pain points from user-generated social media posts using machine learning. Journal of Interactive Marketing 57, 517–539.

Salminen, J., Rao, R.G., Jung, S.g., Chowdhury, S.A., Jansen, B.J., 2020. Enriching social media personas with personality traits: A deep learning approach using the big five classes, in: Artificial Intelligence in HCI: First International Conference, AI-HCI 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings 22, Springer. pp. 101–120.

Sang, J., Lu, D., Xu, C., 2015. A probabilistic framework for temporal user modeling on microblogs, in: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, pp. 961–970.

Sanz-Cruzado, J., Castells, P., Macdonald, C., Ounis, I., 2020. Effective contact recommendation in social networks by adaptation of information retrieval models. Inf. Process. Manag. 57, 102285. URL: https://doi.org/10.1016/j.ipm.2020.102285, doi:10.1016/J.IPM.2020.102285.

Saraswat, M., Chakraverty, S., 2022. Enriching topic coherence on reviews for cross-domain recommendation. The Computer Journal 65, 80–90.

Shao, M.M., Jiang, W.J., Wu, J., Shi, Y.Q., Yum, T., Zhang, J., 2022. Improving friend recommendation for online learning with fine-grained evolving interest. Journal of Computer Science and Technology 37, 1444–1463.

Shen, W., Wang, J., Luo, P., Wang, M., 2013. Linking named entities in tweets with knowledge base via user interest modeling, in: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 68–76.

Song, S., Li, Q., Zheng, X., 2012. Detecting popular topics in micro-blogging based on a user interest-based model, in: The 2012 International Joint Conference on Neural Networks (IJCNN), IEEE. pp. 1–8.

Spasojevic, N., Yan, J., Rao, A., Bhattacharyya, P., 2014. Lasta: Large scale topic assignment on multiple social networks, in: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1809–1818.

Trigka, M., Kanavos, A., Dritsas, E., Vonitsanos, G., Mylonas, P., 2022. The predictive power of a twitter user's profile on cryptocurrency popularity. Big Data and Cognitive Computing 6, 59.

Trikha, A.K., Zarrinkalam, F., Bagheri, E., 2018. Topic-association mining for user interest detection, in: Advances in Information Retrieval: 40th European Conference on IR Research, ECIR 2018, Grenoble, France, March 26-29, 2018, Proceedings 40, Springer. pp. 665–671.

Tupes, E.C., Christal, R.E., 1992. Recurrent personality factors based on trait ratings. Journal of personality 60, 225–251.

Wang, H., Huang, X., Li, L., 2018. Microblog oriented interest extraction with both content and network structure. Intelligent Data Analysis 22, 515–532.

Wang, J., Zhao, W.X., He, Y., Li, X., 2014. Infer user interests via link structure regularization. ACM Transactions on Intelligent Systems and Technology (TIST) 5, 1–22.

Wang, T., Liu, H., He, J., Du, X., 2013. Mining user interests from information sharing behaviors in social media, in: Advances in Knowledge Discovery and Data Mining: 17th Pacific-Asia Conference, PAKDD 2013, Gold Coast, Australia, April 14-17, 2013, Proceedings, Part II 17, Springer. pp. 85–98.

Wang, Y., Ding, S., Xu, X., Jia, W., 2019. The multi-tag semantic correlation used for micro-blog user interest modeling. Engineering Applications of Artificial Intelligence 85, 765–772.

Welch, M.J., Schonfeld, U., He, D., Cho, J., 2011. Topical semantics of twitter links, in: Proceedings of the fourth ACM international conference on Web search and data mining, pp. 327–336.

Wen, Z., Lin, C.Y., 2011. Improving user interest inference from social neighbors, in: Proceedings of the 20th ACM international conference on Information and knowledge management, pp. 1001–1006.

Wu, Q., Burges, C.J., Svore, K.M., Gao, J., 2010. Adapting boosting for information retrieval measures. Information Retrieval 13, 254–270.

Xu, J., Li, H., 2007. Adarank: a boosting algorithm for information retrieval, in: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 391–398.

Xu, J., Lu, T.C., 2015. Toward precise user-topic alignment in online social media, in: 2015 IEEE International Conference on Big Data (Big Data), IEEE. pp. 767–775.

Yalcin, E., Bilge, A., 2022. Evaluating unfairness of popularity bias in recommender systems: A comprehensive user-centric analysis. Inf. Process. Manag. 59, 103100. URL: https://doi.org/10.1016/j.ipm.2022.103100, doi:10.1016/J.IPM.2022.103100.

Yang, L., Sun, T., Zhang, M., Mei, Q., 2012. We know what@ you# tag: does the dual role affect hashtag adoption?, in: Proceedings of the 21st international conference on World Wide Web, pp. 261–270.

Yin, C., Zhang, X., Liu, L., 2020. Reposting negative information on microblogs: Do personality traits matter? Information Processing & Management 57, 102106.

Yin, H., Cui, B., Chen, L., Hu, Z., Zhou, X., 2015. Dynamic user modeling in social media systems. ACM Transactions on Information Systems (TOIS) 33, 1–44.

Yoon, H.J., Tourassi, G., 2014. Analysis of online social networks to understand information sharing behaviors through social cognitive theory, in: Proceedings of the 2014 Biomedical Sciences and Engineering Conference, IEEE. pp. 1–4.

Yu, W.J., Hung, S.Y., Yu, A.P.I., Hung, Y.L., 2023. Understanding consumers' continuance intention of social shopping and social media participation: The perspective of friends on social media. Information & Management , 103808.

Zarrinkalam, F., Fani, H., Bagheri, E., Kahani, M., Du, W., 2015. Semantics-enabled user interest detection from twitter, in: 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), IEEE. pp. 469–476.

Zarrinkalam, F., Faralli, S., Piao, G., Bagheri, E., et al., 2020. Extracting, mining and predicting users' interests from social media. Foundations and Trends® in Information Retrieval 14, 445–617.

Zarrinkalam, F., Kahani, M., Bagheri, E., 2018. Mining user interests over active topics on social networks. Information Processing & Management 54, 339–357.

Zarrinkalam, F., Kahani, M., Bagheri, E., 2019. User interest prediction over future unobserved topics on social networks. Information Retrieval Journal 22, 93–128.

Zhao, W.X., Jiang, J., Weng, J., He, J., Lim, E.P., Yan, H., Li, X., 2011. Comparing twitter and traditional media using topic models, in: Advances in Information Retrieval: 33rd European Conference on IR Research, ECIR 2011, Dublin, Ireland, April 18-21, 2011. Proceedings 33, Springer. pp. 338–349.

Zhao, Z., Cheng, Z., Hong, L., Chi, E.H., 2015. Improving user topic interest profiles by behavior factorization, in: Proceedings of the 24th International Conference on World Wide Web, pp. 1406–1416.

Zheng, G., Zhang, F., Zheng, Z., Xiang, Y., Yuan, N.J., Xie, X., Li, Z., 2018. Drn: A deep reinforcement learning framework for news recommendation, in: Proceedings of the 2018 world wide web conference, pp. 167–176.

Zheng, W., Ge, B., Wang, C., 2019. Building a tin-lda model for mining microblog users' interest. IEEE Access 7, 21795–21806.

Zheng, X., An, D., Chen, X., Guo, W., 2016. Interest prediction in social networks based on markov chain modeling on clustered users. Concurrency and Computation: Practice and Experience 28, 3895–3909.

Zhu, Y., Zhong, E., Pan, S.J., Wang, X., Zhou, M., Yang, Q., 2013. Predicting user activity level in social networks, in: Proceedings of the 22nd ACM international conference on Information & Knowledge Management, pp. 159–168.

Gil de Zúñiga, H., Diehl, T., Huber, B., Liu, J., 2017. Personality traits and social media use in 20 countries: How personality relates to frequency of social media use, social media news use, and social media use for social interaction. Cyberpsychology, Behavior, and Social Networking 20, 540–552.