# Temporal Conformity-aware Hawkes Graph Network for Recommendations

Chenglong Ma
chenglong.ma@student.rmit.edu.au
Royal Melbourne Institute of Technology
Melbourne, Australia

Yongli Ren
yongli.ren@rmit.edu.au
Royal Melbourne Institute of Technology
Melbourne, Australia

Pablo Castells
pablo.castells@uam.es
Universidad Autónoma de Madrid
Madrid, Spain

Mark Sanderson
mark.sanderson@rmit.edu.au
Royal Melbourne Institute of Technology
Melbourne, Australia

## ABSTRACT

Many existing recommender systems (RSs) assume user behavior is governed solely by their interests. However, the peer effect often influences individual decision-making, which leads to conformity behavior. Conventional solutions that eliminate indiscriminately such bias may cause RSs to neglect valuable information and depersonalize the recommendation results. Also, conformity can transform into user interest, e.g., discovering new tastes after a glance at popular music. By better representing different forms of conformity influence, we can do a better job at interest mining and debiasing. In certain extreme circumstances, the herd effect may be exacerbated by user anxiety with uncertainty (e.g., panic buying during the COVID-19 pandemic). RSs may thus fail to respond in time due to sudden and dramatic changes. Moreover, many existing studies potentially conflate conformity bias with popularity bias and lump together various factors responsible for differences in popularity. In this paper, we identify two distinct types of conformity behavior: informational conformity and normative conformity. To address this, we introduce the TCHN model, which utilizes attentional Hawkes processes to disentangle user self-interest and conformity in a personalized manner. Our approach incorporates temporal graph attention networks to capture users' stable and volatile dynamics. We conduct experiments on three real-world datasets, which uncover diverse levels of conformity among users. The results show that TCHN excels in recommendation accuracy, diversity, and fairness across various user groups. [1]

## CCS CONCEPTS

• **Information systems** → **Recommender systems**.

## KEYWORDS

User Interest, Conformity, Temporal Graph Attention Network, Hawkes Process, Recommendations

---

[1]Code is available at https://github.com/ChenglongMa/TCHN

## 1 INTRODUCTION

Recommender systems utilize user behavioral data to capture user profiles and operate under the assumption that an individual's actions are intrinsically motivated by their interests [49]. However, in an informational cascade [5], people can frequently base their decisions not only on their own information but also on the influence of their peers within the same social community [39, 53]. Moreover, in different situations, people will actively shift the importance of the two factors to alleviate cognitive dissonance in their decision-making process [2, 4, 43]. For instance, when choosing a music album, users are more likely to rely on their own sensory experience than the opinions of others [47]; however, during extreme events such as the COVID-19 pandemic, the Fear of Missing Out (FOMO) phenomenon [44] comes into play, and people's purchasing behavior can be significantly influenced by others (e.g., panic buying in COVID-19 pandemic) [38, 68]. Due to the aforementioned fundamental assumption, traditional RSs are known to introduce popularity and conformity biases [10, 71, 72], where the learning process fails to differentiate between a user's genuine interest and the biased behavior.

Despite numerous studies [6, 33, 46, 48, 61, 71, 72] dedicated to this issue, several challenges remain to be addressed. First, most of the existing research views conformity behavior negatively as an interfering factor in user profiling [6, 46, 48]. However, we argue that conformity behavior reflects the combination of individuals' cognition of their inner needs [1, 39–41] and their perception of external circumstances [4, 43, 53, 68]. Eliminating behavioral data indiscriminately can result in the loss of valuable information and eventually lead to poor recommendations. Thus, we argue that a personalized method is needed to handle people's inner interest and the conformity bias in the system for each individual. While it is widely agreed that conformity causes users to blindly focus on popular items [33, 61, 71, 72], we may be overlooking individual differences in users' attitudes toward such items based on their personalities. For example, some users, known as "deep divers" in this research, have an unshakable interest in niche topics less influenced

by environmental changes or others' opinions. While strategies aimed at eliminating biases may benefit these "deep divers", they may have a detrimental effect on another group of people who tend to prefer popular items and follow trends, denoted as "surfers" here.

Second, conformity bias and popularity bias are not equivalent. As the former is hardly observed and measured explicitly, many existing studies, e.g., [6, 61, 71, 72], have resorted to using *popularity* to characterize conformity. However, we argue that conformity behavior is only a sufficient but not necessary condition for popularity. The latter has its own complicated causes, such as the high quality of a product [7]. Individuals conform to peers either due to a lack of relevant knowledge (informational conformity) or to avoid isolation (normative conformity) [14, 16, 30]. This aligns with the fundamental logic of RSs, which is to intervene only when users' interests and behaviors do not match.

Third, these studies treat conformity bias as a *static* factor interfering with the operation of RSs. They learn conformity behavior inflexibly and disregard its temporal sensitivity and volatility. A potential interchange exists between user interest and conformity behavior: spontaneous conformity may help users develop new interests, whereas their enthusiasm for one item may potentially fade over time when the external stimulus wears off [4, 68]. Therefore, it is crucial to consider the temporal dimension of the relationship and how it may evolve over time.

Finally, the intensity and scale of conformity matter due to the fragility of information cascades [5, 22]. For example, some extreme outlier events (e.g., pandemic-like events [38, 39]) can cause abrupt population-scale concept drift in cascades. It further aggravates popularity bias [39] and out-of-distribution (OOD) issues [25] in RSs. Moreover, the population-scale concept drift might lead individuals to rush to products that normally go unnoticed, e.g., hand sanitizer during COVID-19. Conventional RSs are often unable to adapt to the changes, resulting in failure to balance recommendation accuracy and diversity [38, 39].

To tackle these challenges, we propose a **T**emporal **C**onformity-aware **H**awkes **N**etwork (TCHN), which decouples user behavior patterns in RSs as a mutual excited Hawkes process combining user interest and conformity. It draws inspiration from both Maslow's hierarchy of needs [40, 41] and social identity theory [53], which posit that individuals' decisions are influenced by their own inner needs as well as the opinions of others in their social environment:

$$D = \theta I + (1 - \theta)C, \tag{1}$$

where $D$, $I$ and $C$ represent "decision", "interest" and "conformity" respectively; $\theta$ is used to balance the influence of the two. It is worth noting that the motivation behind user behavior is a complex and difficult-to-observe sociological issue. In this study, we have only focused on the principal factors related to recommendation behavior. We disentangle their representations at the scale of the user-item interaction sequence graph. The resulting sequential embeddings are fed into customized attention-based Hawkes processes for preference prediction. TCHN treats user interests as stable information and conformity as a volatile and time-limited signal. Both user interest and conformity guide how to make a new decision, but only long-lasting user interest is stored in user representation. It dissociates the influence of user conformity behavior and personalizes its effect in the latest user profiling.

To sum up, the contributions of our research are as follows:

- We introduce a personalized attention graph network based on the Hawkes process to untangle user decision-making into distinct components of interest and conformity. Our model adeptly incorporates user conformity behavior in a personalized manner, avoiding simplistic bias removal.
- Dynamic modeling of two forms of conformity behavior enables the model to distinguish and uncover deep user interests during interactions.
- Extensive experiments on three real-world datasets, featuring diverse scales of conformity behavior, show that our model enhances recommendation accuracy and diversity for both "surfers" and "deep divers".

## 2 RELATED WORK

### 2.1 Conformity in User Interactions

Conformity, characterized by adherence to group norms, involves individuals relying on collective wisdom for decision-making, even if it contradicts their personal beliefs [13, 53]. In RSs, this manifests as users opting for popular items over those aligned with their genuine interests. Besides, since the intrinsic motivation of conformity cannot be directly observed, this issue is often referred to as a *popularity bias*.
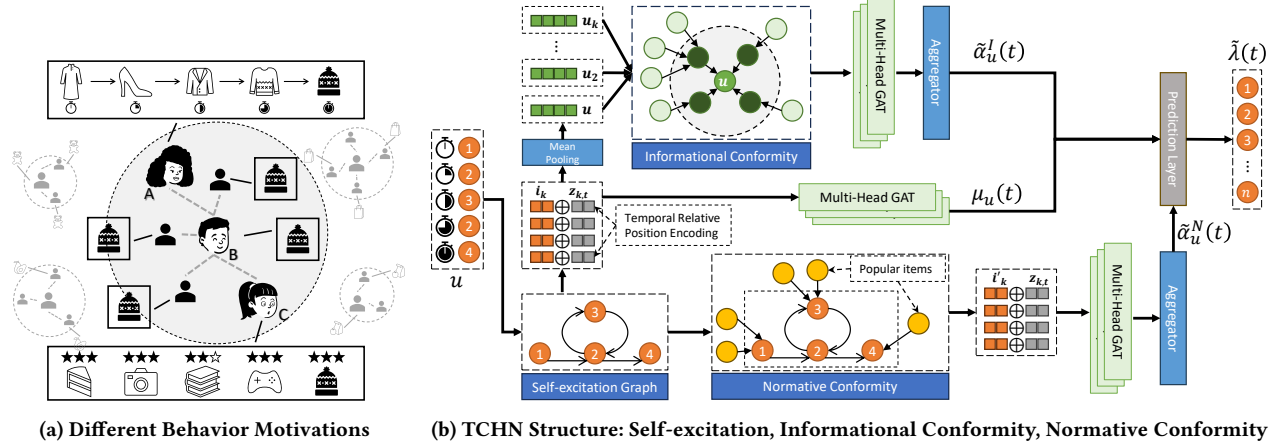
Various studies, such as [48] and [46], address popularity bias through inverse propensity scoring (IPS) methods. Bonner and Vasile [6] guides biased training using unbiased intervention data, while Zhang et al. [71] considers item popularity as a valuable indicator for learning user preferences. However, their approach treats popularity uniformly across all users, neglecting individual differences in preference. Zheng et al. [72] views popularity as an observable outcome of conformity and employs it to disentangle user decisions into interests and conformity, drawing on causal mechanisms [26, 28]. However, discerning underlying motivations solely from observed data remains challenging. We contend that asserting conformity as a necessity for popularity is arbitrary.

Deutsch and Gerard [16] delineates two forms of conformity: informational and normative conformity. Informational conformity stems from the belief in the group's superior knowledge, while normative conformity arises from a desire for social approval or fitting in [14]. Although conformity can sometimes be solely based on information or norms, it usually, especially in RSs, involves both factors simultaneously [30]. Thus, leveraging conformity should be gentle and intervention warranted only when deviations in interest and behavior occur.

### 2.2 Sequential Recommendations

Information dissemination in user interactions with RSs is an aggregation of user needs and the influence of preceding peer behaviors [5]. While conformity is often considered contrary to user interests, it can turn into user interest under certain circumstances.

Sequential RSs model user temporal dynamics from historical interactions. Recently, recurrent neural networks (RNNs) have gained popularity for sequential recommendations [17, 52, 73]. GRUs, as typical RNN implementations, model user interactions by maintaining a memory of past inputs. Zhou et al. [73] proposes a two-layer

(a) Different Behavior Motivations

(b) TCHN Structure: Self-excitation, Informational Conformity, Normative Conformity

GRU-based network to extract and model the evolution of users' deep interests. Transformer models, another prevalent class of sequential neural network models, have achieved state-of-the-art performance in recommendation tasks [11, 52]. Attention mechanisms [55] enable models to focus on specific parts of the input depending on target attributes. Ying et al. [67] propose a hierarchical attention network for next-item recommendation tasks that balances user long-term preference evolution and new characteristics in their short-term preference.

## 2.3 Hawkes Process

Conventional sequential models effectively capture ordinal dependencies between user behaviors but potentially neglect the time interval between two successive actions [35]. Temporal point processes (TPP) [15, 18] are mathematical abstractions of stochastic counting processes for realizations of asynchronous events occurring in continuous time. Hawkes process (HP), a self-exciting variant of TPP, assumes the arrival of an event stimulates subsequent events in the near future with various intensities [24]. HP's mutual excitation property facilitates user influence inference in social interactions. Recently, concerns have arisen about misspecifications in HP models due to pre-assumed intensity functions that do not fit the law of events [66]. Many researchers attempt to adapt sequential neural networks to TPP learning. Mei and Eisner [42] extends HP by a long short-term memory (LSTM) network that relaxes the excitation constraint so as to simulate elevated or inhibited offspring events. Zhang et al. [70], Zuo et al. [74] propose attention-based HPs to capture deep timing dependencies between events and alleviate the gradient explosion and vanishing issues in RNNs. Additionally, numerous studies integrate HP models into RSs to tackle various challenges, e.g., preference evolution [3], repeat consumption modeling [57], adaptability in user long/short-term preferences [58].

## 2.4 Graph Representation Learning

To study the mutual excitation of user interaction events in Hawkes processes, understanding the topological structure of information diffusion is essential. RSs, especially collaborative filtering (CF) models, often exhibit properties of social networks, allowing effortless conversion of user interactions into a graph structure. Utilizing graph neural network (GNN) techniques, such as the gated graph sequence neural network (GGNN) [36] and the graph attention network (GAT) [56], in sequential RSs enables the learning of deep representations of user nodes in the social graph and understanding the influence of neighboring nodes on decision-making in information cascades. However, the *permutation-invariant* aggregation function used by GGNN and its subsequent models in this paradigm [9, 60] during message passing can potentially result in the loss of valuable ordering information within the neighborhood [12]. Attention mechanisms used in GATs allow the network to selectively attend to relevant neighbors and aggregate information from them while ignoring irrelevant nodes [56, 59, 62].

## 3 METHODOLOGY

Prior to elaborating on our approach, we aim to clarify the problem through the visual aid of Fig. 1a. The figure encapsulates three distinct user archetypes interacting with the hat, each driven by unique motivations. User $A$, characterized as a "deep diver", exhibits a self-motivated interaction with the hat (i.e., self-excitation); In contrast, user $B$ consults information from neighboring nodes with shared interests, culminating in the selection of the hat (i.e., informational conformity). Meanwhile, user $C$, typified as a "surfer", demonstrates a propensity for emulating peers and pursuing trending items (i.e., normative conformity). Our research endeavor is dedicated to a solution capable of discerning and accommodating these heterogeneous motivations to provide adaptive recommendations.

## 3.1 Problem Formulation

Given a user set $\mathcal{U}$ ($M = |\mathcal{U}|$) and an item set $\mathcal{I}$ ($N = |\mathcal{I}|$), let $S_u(t) = [s_1, s_2, \cdots, s_k]$, with $s_k = (i_t, r_{i,t}, t)$, be a historical sequence of length $k$ for user $u$ and her recent $k$ engagements with the item set $\mathcal{I}$ up to time $t$, where $r_{i,t}$ is the feedback[2] of item $i$ at time $t$ given by user $u$. To elaborate further, we augment available data by reframing the historical sequence into a sequence matrix. This matrix includes supplementary paired samples structured as

---

[2]w.l.o.g, the feedback can be explicit (e.g., ratings) or implicit (e.g., click).

$\left[ \langle S_u(t), i_{t+1}^+, i_{t+1}^- \rangle \right]_{t=1}^{k-1}$. Here, $i_{t+1}^+$ is the next item that user $u$ interacts with after $S_u(t)$ while $i_{t+1}^-$ is a hypothetical negative or unobserved instance for the same user. This allows us to capture the sequential nature of user behavior and model the evolution of user preferences over time [54].

Our final goal is to recommend a set of items for user $u$ that strike a balance between accuracy and diversity. Here, we present a Hawkes process based graph attention network TCHN (as shown in Fig. 1b) to address the problem of temporal evolution in user interests and conformity, which includes the following four main components: 1) **self-excitation** component to extract deep interest from users' own interaction history; 2) **informational conformity** modeling to aggregate influence of users' neighboring branches in an information cascade; 3) **normative conformity** modeling to measure the acceptance level of individuals on popular fashion; 4) **recommendation generation** component to fuse the three information signals to produce the final personalized recommendations. Before diving into the details of the model, we first introduce the preliminaries and the temporal relative position encoding method.

## 3.2 Preliminaries

*3.2.1 Hawkes Process* From a point process perspective, we treat the interaction $(s_t)$ given by a user on an item as an *event* and the sequence involving the same user $(S_u(t))$ as a realization of a Hawkes process. Formally, in an infinitesimal time window $[t, t + \Delta t)$, the conditional intensity function $\lambda^*(t)$ [66] is:

$$\lambda^*(t) = \lambda(t \mid S_u(t)) = \lim_{\Delta t \to 0} \frac{\mathbb{E}[\Delta N(t) \mid S_u(t)]}{\Delta t}, \quad (2)$$

where $\mathbb{E}[\Delta N(t) \mid S_u(t)]$ is the expected number of events occurred in $(t+\Delta t]$ conditional on $S_u(t)$. Also, we assume two events coincide with probability 0, i.e., $\Delta N(t) \in \{0, 1\}$ [45]. The intensity $\lambda^*(t)$ denotes the occurrence rate of future events.

In this paper, we season the intensity function with the mutual excitation property of the Hawkes process. Also, we define an $M$-dimensional Hawkes process for all user interaction sequences where a single event in one dimension can excite or inhibit the intensities of all dimensions. The intensity function of $u$-th dimension can be updated as:

$$\lambda_u(t) = \mu_u(t) + \sum_{v \in U} \sum_{t_{v,i} < t} \alpha_{u,v}(t) \kappa_{u,v}(t - t_{v,i}), \quad (3)$$

where $\mu_u(t)$ is the base intensity of user $u$ at time $t$ triggered by their own intent. For the process, this arrival is spontaneous and independent of preceding events (a.k.a exogenous intensity). $t_{v,i}$ is the time when user $v$ interacted with item $i$. $\alpha_{u,v}(t)$ describes the strength of influence that user $v$ exerts on user $u$. Note that $\alpha_{u,v}(t)$ is unidirectional and not symmetrical, i.e., $\alpha_{u,v}(t) \neq \alpha_{v,u}(t)$. $\kappa_{u,v}(t - t_{v,i})$ is the triggering kernel that quantifies the rate of occurrence after the realization $t_{v,i}$. The most commonly used form of this kernel is the exponential kernel function: $\kappa_{u,v}(\Delta t) = e^{-\beta_{u,v}(\Delta t)}$, where $\beta_{u,v}$ controls the decay rate. The second addend in Eq. (3) captures the aggregated influence of other point processes (a.k.a endogenous intensity). A typical HP [24] only supports positive excitation, but here we relax such constraint and enable the inhibition effect [42, 58, 74], i.e., $\alpha_{u,v}(t), \mu_u(t) \in \mathbb{R}$.

In the context of RSs, the intensity function $\lambda_u(t)$ neatly crystallizes the user's decision-making process in Eq. (1). Users' intrinsic interest activates their base intensity; meanwhile, preceding events (interactions) from other individuals excite (or inhibit) the target user as a reference signal.

## 3.3 Embedding Representation

In our proposed model, we define three key entities within the latent embedding space: users, items, and time. To begin, the model initially maps one-hot representations of items to a unified low-dimensional embedding space, denoted as $i \in \mathcal{I} \mapsto \mathbf{i} \in \mathbb{R}^d$, where $d$ is the embedding dimension, and it is constrained such that $d \ll \min\{M, N\}$. Further, given two timestamps $\langle t_a, t_b \rangle$, we translate their time difference into a temporal relative positional embedding, i.e., $|t_b - t_a| \mapsto \mathbf{z}_{a,b}$. Consequently, the interaction sequence $S_u(t)$ can be encoded as latent vectors $S_u(t) \mapsto \mathbf{S}_u(t) = \left[ \Theta\left(\mathbf{i}_\tau, \mathbf{z}_{\tau,t}\right) \right]_{\tau=1}^t$, where $\Theta(\cdot)$ is a differentiable, permutation-invariant function, such as element-wise summation or concatenation.

Remarkably, we do not explicitly encode user profiles, but instead aggregate their historical interaction sequences $S_u(t)$ to serve as user representations. Formally, user $u$ is mapped to the embedding $\mathbf{u} := \Phi(\mathbf{S}_u(t))$, where $\Phi(\cdot)$ is a differentiable function. The implementation details of $\Theta(\cdot)$ and $\Phi(\cdot)$ will be discussed in subsequent sections.

## 3.4 Temporal Relative Position Encoding

As outlined in Sec. 3.2.1, our approach involves scrutinizing the user's interaction history within the Hawkes point process framework. This methodology entails observing how new interactive events can either excite or inhibit the intensity of ongoing events, with these effects gradually diminishing over time according to the specified kernel function $\kappa_{u,v}(\Delta t)$ in Eq. (3). In contrast to RNN networks, standard attention-based transformer models lack the capability to discern position differences in the input sequence [50, 55]. Consequently, our learning model necessitates the inclusion of supplementary positional representations within the input data. Traditional order-position encoding [55] or absolute value of time cannot disclose rigorous time transition information as effectively as relative timespan [65]. In this work, we revamp the approach proposed by [50, 65] into our model to expose the pairwise temporal relationships between different interactions. Formally, we first establish a positive semi-definite (PSD) temporal kernel $\kappa(t_a, t_b)$ that complies with Bochner's theorem assumption (proof is supplied in Appx. A). Then, the kernel value is quantized to an integer, serving as the relative position for events at time $t_a$ and $t_b$:

$$\kappa(t_a, t_b) = \theta_z \ln\left(\beta_{a,b}\left(|t_b - t_a|\right) + 1\right) \quad (4)$$

$$z_{a,b} = \lceil \kappa(t_a, t_b) \rceil, \quad (5)$$

where $\theta_z > 0$ is a hyper-parameter to manage the granularity and capacity for the positional embedding space. $\beta_{a,b} > 0$ governs the decay rate for different events. We stipulate that $\theta_z = 10$ and the time unit of $t_a$ and $t_b$ are *days* in our experiment settings. Additionally, we postulate that different individuals share the same sensitivity to time differences for simplicity; hence, we set $\beta_{a,b} = \beta = 1$. This assumption implies that recent events exert a more

subtle influence on their immediate decisions, while events more distant from the present have a diminishing and less pronounced impact. Finally, we encode the relative positional information into a position-embedding matrix, where $z_{a,b} \in \mathbb{R}^+ \mapsto \mathbf{z}_{a,b} \in \mathbb{R}^d$.

In the following sections, we will provide detailed insights into the various components of our methodology as integrated within the context of the Hawkes process and a graph neural network framework.

## 3.5 Self-excitation Graph

Intuitively, the base intensity $\mu_u(t)$ of a user on a specific item at time $t$ is primarily determined by their intrinsic interest. In addition, individuals' interests are subject to change over time as their inner needs evolve [31, 39]. We derive nutritional value from the user's historical interaction sequence. Additionally, more recent interactions carry greater weight in determining the user's current interests and needs.

To implement this concept, we transform each interaction sequence $S_u(t)$ into a user-specific graph. This enables the acquisition of item-transition patterns within the sequence $S_u(t)$ by utilizing a time-aware graph attention network (TGAT). Given $S_u(t)$, let $G_u(t) = (V_u(t), E_u(t))$ represent a directed user-specific graph, where $V_u(t) \subset I$ denotes the set of interacted items in $S_u(t)$, and $E_u(t)$ is the edge set. Each edge connects two adjacent items $(i_a, i_b)$ in $S_u(t)$ and points from $i_a$ to $i_b$ if $t_b > t_a$. Next we introduce TGAT.

### 3.5.1 Time-aware Graph Attention Layer
Analogous to GAT [56], the TGAT layer can be conceived as a local aggregation operator. It takes as input the hidden representations and temporal relative position embeddings of the neighborhood surrounding the target node. The output is the time-aware representation for the target node at the current time $t$.

As mentioned in Sec. 3.3, given current timestamp $t$, the hidden representation of $G_u(t)$ is specified by

$$\mathbf{H}_u(t) = \Theta\left(\mathbf{I}_u(t), \mathbf{Z}_u(t)\right)^\top = [\mathbf{i}_{t_1} + \mathbf{z}_{t_1,t}, \cdots, \mathbf{i}_{t_k} + \mathbf{z}_{t_k,t}]^\top \quad (6)$$

where $\mathbf{I}_u(t) = [\mathbf{i}_{t_k}]_{i \in G_u(t)}$ is the node embedding matrix of $G_u(t)$, and $\mathbf{Z}_u(t) = [\mathbf{z}_{t_k,t}]_{t_k \in G_u(t)}$ is the temporal embedding matrix of relative positions between nodes in $G_u(t)$. $\Theta(\cdot)$ is chosen as element-wise summation.

Subsequently, we pass $\mathbf{H}_u(t)$ and the embedding of the target item $\mathbf{i}_t$ into the self-attention module. And the attention output will be treated as the base intensity of user $u$ on item $i$:

$$\tilde{\mu}_{u,i}(t) = Attention(\mathbf{H}_u(t), \mathbf{i}_t) = \text{softmax}\left(\frac{\mathbf{H}_u(t)\mathbf{i}_t^\top}{\sqrt{d}}\right) \quad (7)$$

Finally, Eq. (7) models the evolution of the interest of user $u$, and captures her intrinsic preference on the target item $i$.

## 3.6 Conformity Graphs

The classical Hawkes process inherently captures the strength of unidirectional influence from the predecessors to the target user, i.e., $\alpha_{u,v}(t)$ in Eq. (3). However, we speculate that this influence entangles the conformity of individuals. In other words, interactions between individuals are impacted not only by their shared preferences but also by peer effects. We extend the classical model

to disentangle various forms of conformity, namely informational conformity and normative conformity [16]. As aforementioned, the two forms of conformity often occur concurrently [30], and their influences usually vary given different users and items. Therefore, inspired by [34], we disentangle the influence strength $\alpha_{u,v}(t)$ into two additive terms:

$$\alpha_{u,v}(t) = \theta_{u,v}^I(t)\alpha_{u,v}^I(t) + \theta_{u,v}^N(t)\alpha_{u,v}^N(t), \quad (8)$$

where $\alpha_{u,v}^I(t)$ and $\alpha_{u,v}^N(t)$ are the two forms of conformity respectively. The time-varying coefficients $\theta_{u,v}^I(t)$ and $\theta_{u,v}^N(t)$ balance their contributions at time $t$. The intensity function of the *conformity-aware Hawkes process* can be updated as follows by substituting it to Eq. (3):

$$\lambda_u(t) = \mu_u(t) + \sum_{v \in U} \sum_{t_{v,i} < t} \left(\theta_{u,v}^I(t)\alpha_{u,v}^I(t) + \theta_{u,v}^N(t)\alpha_{u,v}^N(t)\right) \kappa_{u,v}(t - t_{v,i})$$
$$(9)$$

Next, we describe how to quantify the two forms of conformity in RSs. Before that, we need to establish a clear scope and definition for both within the framework of RSs.

### 3.6.1 Informational Conformity
In the context of RSs, informational conformity could manifest itself when users rely on the recommendations of the system because they believe that the system has more information and knowledge about the items being recommended. Formally, we define **informational conformity** as the extent to which users adopt the aggregated preferences of other *like-minded users* (neighbors) in the system [49].

We formulate the modeling of informational conformity in a GNN, allowing for effective learning of mutual influences between nodes. Given the user set $\mathcal{U}$, we build a conformity graph based on the social connections between users. Without loss of generality, in the framework of RSs, we infer the social connection between users through the interaction between users and items instead of introducing additional social information, such as social actions (e.g., following and liking) [62]. This is because merely understanding the structure of a social action-based topology is inadequate to tackle this issue, as the connections between individuals do not always imply the occurrence of social interactions among them, whereas they can be affected by some strangers unintentionally [34]. For example, suppose both users $u$ and $v$ interacted with the same item $i$ at time $t_u$ and $t_v$. We build an edge pointing from $u$ to $v$ if $t_v > t_u$ that indicates user $u$'s action can potentially affect (excite or inhibit) user $v$'s decision even though they do not know each other. Thus the social network is formulated as a user-user directed graph: $G = \langle \mathcal{U}, E \rangle$, where $E$ is the set of social connections among $\mathcal{U}$. Particularly, the weight of the edge between $u$ and $v$ indicates the social influence of user $u$ on user $v$. According to social psychology theories [2, 16, 69], the degree of influence exerted by user $u$ positively correlates with the level of informational conformity exhibited by user $v$ towards user $u$. Based on this, it can be inferred that if individuals $v$ and $u$ interact *frequently*, it is advisable to enhance their informational influence. Additionally, we hypothesize that user $v$ will likely conform to user $u$ when user $v$ consistently agrees with user $u$ (*opinion polarity* w.r.t an item). Thus, given a historical interaction sequence $S_u(t)$ (resp. $S_v(t)$) of user $u$ (resp. $v$) the weight of edge $u \to v$ can be quantified as:

$$w_{u,v} = \frac{\sum_{(r,t) \in \Lambda(u,v)} (r_u - \bar{r}_u)(r_v - \bar{r}_v) \exp\left(-\beta_{u,v}(t_v - t_u)\right)}{\sqrt{\sum (r_u - \bar{r}_u)^2 \sum (r_v - \bar{r}_v)^2}} \quad (10)$$

where $\Lambda(u,v) = \{S_u(t) \cap S_v(t) \mid i_u = i_v \wedge t_v >_u\}$ is the intersection of ratings of items interacted by both $u$ and $v$; meanwhile, $t_v > t_u$ in each interaction. Besides, in line with the definition of the Hawkes process in Eq. (3), we penalize the opinion polarity far from the present by the decay function. That is, the weight reveals how likely it is that the decision of user $v$ is infected by user $u$ in the presence of informational conformity.

We apply a GAT on the social influence graph to learn about the informational conformity strength in the ego social network of the target user. The GAT layer utilizes a multi-head masked attention mechanism to aggregate the influence of neighboring nodes on the target node. Firstly, the attention coefficient between the target node $v$ and one of its neighboring nodes $u$ at the $k$-th head is:

$$\mathbf{h}_u = \Phi(\mathbf{H}_u(t)) = \frac{1}{k} \sum (\mathbf{i}_{t_k} + \mathbf{z}_{t_k}), \mathbf{h}_v = \Phi(\mathbf{H}_v(t)) \quad (11)$$

$$c_{u,v}^k = \phi\left(\left[w_{u,v}(\mathbf{W}_k \cdot \mathbf{h}_u) || (\mathbf{W}_k \cdot \mathbf{h}_v)\right]\right), \quad (12)$$

$$\alpha_{u,v}^k = Attention\left(c_{u,v}^k\right) = \text{softmax}\left(\frac{c_{u,v}^k}{\sqrt{d}}\right), \quad (13)$$

where $\phi(\cdot)$ is the LeakyReLU activation function, $w_{u,v}$ is the weight of edge $u \rightarrow v$, $\mathbf{W}_k$ is the shared parameter matrix and $\mathbf{h}_u$) (resp. $\mathbf{h}_v$)) is the node embedding of $u$ and $v$, which is aggregated from self-excitation graph $G_u(t)$; $[\cdot||\cdot]$ is the concatenation operation of embeddings.

Its node representation is computed by aggregating the influence from the neighboring nodes:

$$\mathbf{h}_v'(K) = \frac{1}{K} \sum_{k=1}^{K} \sigma\left(\sum_{u \in \mathcal{N}_v} \alpha_{u,v}^k \mathbf{W}_k \mathbf{h}_u\right), \quad (14)$$

where $\sigma(\cdot)$ is the Sigmoid activation function. We apply a *mean-pooling* operation to fuse the $K$-head output. The informational influence strength will be:

$$\tilde{\alpha}_{u,v}^I(t) = \frac{1}{K} \sum_{k=1}^{K} \alpha_{u,v}^k. \quad (15)$$

*3.6.2 Normative Conformity* Normative conformity refers to the phenomenon in which individuals unconsciously conform to the norms of a particular social group, suppressing their rational thinking [53]. In the context of RSs, this is evident when users change their usual behavior habits and preferences to align with those of the crowd (e.g., panic buying [68]), or adjust their own evaluations to adhere to the average opinion [20]. Formally, we define **normative conformity** as the extent to which users adopt the aggregated preference on *recently popular items* within the system. Specifically, we speculate that more popular items and recent interactions can convey stronger conformity signals to the target user. Then, the target user can respond to the signals by shifting their original intentions. For instance, online shopping users refer to the sale promotion on the home page (e.g., "Best sellers" in Amazon [49])

to place an order; music platform users add recent hits to their playlists [47].

Following this definition, to quantify the attractiveness of fashion to a user during the decision-making process, we gather the most popular items $P(t)$ in the system within the recent time window. The width of the time window is denoted as $\tau$, meaning that the interactions considered must have occurred at a time $t$ such that $t \geq (t_d - \tau)$. Here, $t_d$ represents the time at which the user makes a decision or receives a recommendation.

We first map the items in the collection $P(t)$ to their embedding space: $P(t) \in \mathbb{R}^n \mapsto \mathbf{P}(t) \in \mathbb{R}^{n \times d}$, where $n = |P(t)|$. Next, we compute the popularity-weighted attention score, which can be treated as the normative conformity strength:

$$c(\mathbf{P}(t), i) = \mathop{||}\limits_{j=1}^{n} \phi\left(\frac{p_j}{p_i} \cdot \mathbf{W} \cdot \mathbf{j}\right), \quad (16)$$

$$\alpha(c(\mathbf{P}(t), i)) = \text{softmax}\left(\frac{c(\mathbf{P}(t), i)\,\mathbf{i}^\top}{\sqrt{d}}\right), \quad (17)$$

where $\mathop{||}\limits_{j=1}^{n} (\cdot)$ is $n$ concatenation operations, $\phi$ is the ReLu activation function. Importantly, we inject a relative popularity weight to the attention coefficient $c(\mathbf{P}(t), i)$, i.e., $p_j/p_i$, where $p_j$ (resp. $p_i$) is the number of interactions of item $j$ (resp. $i$) within the time window $\tau$. Specifically, the attention score amplifies the impact of more popular items while suppressing the influence of less popular items, which will be treated as the normative influence strength in our model: $\tilde{\alpha}_i^N(t) = \alpha(c(\mathbf{P}(t), i))$.

### 3.7 Recommendation Generation

Through the analysis of the above models and sociological research [2, 13, 69], we believe that the two forms of conformity play different roles in shaping individuals' beliefs and behavior patterns. Specifically, informational conformity has the potential to induce genuine and enduring alterations in beliefs. The result of informational influence is normally private acceptance: the development of the interest of individuals. Instead, normative conformity is less likely to cause lasting change and more of a coping mechanism people use to avoid isolation. Once the external stimulus diminishes, individuals consciously correct this behavior. However, prior to that, this conformity signal plays an important role in influencing people's decision-making.

Therefore, when training the model, we only update the representations of users and items in *self-excitation* and *informational conformity* components. *Normative conformity* only participates in decision-making without affecting the deep representations.

The preference of the target user $u$ on the target item $i$ at time $t$ is quantified as the intensity of the conformity-aware Hawkes process:

$$\tilde{\lambda}_{u,i}(t) = \tilde{\mu}_u(t) + \sum_{v \in U} \theta_{u,v}^I(t)\tilde{\alpha}_{u,v}^I(t) + \theta_{u,v}^N(t)\tilde{\alpha}_i^N(t). \quad (18)$$

L2 regularization is employed on the parameters of the mapping function to address overfitting, although it is not demonstrated here for brevity.

**Table 1: Statistics of Datasets**

| Dataset | #Users | #Items | #Interactions |
|---------|--------|--------|---------------|
| Diginetica | 72,013 | 29,453 | 580,490 |
| Kuai | 7,176 | 10,612 | 1,153,787 |
| Yelp | 78,163 | 57,718 | 1,856,942 |
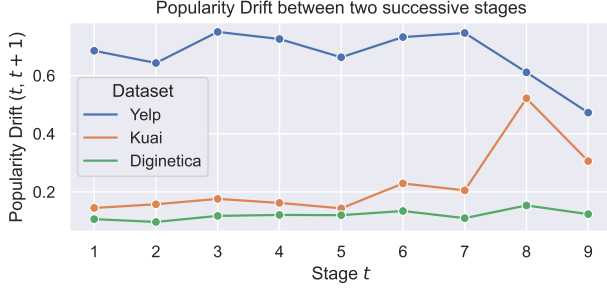


**Figure 2: Popularity Drift between $t$ and $t + 1$**

We apply a contrastive learning scheme to train the model. Given the sequence $\left[ < S_u(t), i^+_{t+1}, i^-_{t+1} > \right]^{k-1}_{t=1}$, the loss function can be computed as:

$$L = -\frac{1}{k} \sum^k_{t=1} \left( y \log \tilde{\lambda}_{u,i^+}(t) + (1-y) \log(1 - \tilde{\lambda}_{u,i^-}(t)) \right) \quad (19)$$

## 4 EXPERIMENTS

### 4.1 Experiment Settings

*4.1.1 Datasets* We evaluate the proposed model on three real-world datasets: Diginetica [3], Kuai [21] and Yelp [4], because: 1) their degrees of sparsity vary and are collected from different domains; 2) their users have different degrees of interest and conformity drift. We map the labels of all positive interactions to "1", and others to "0". Following [63, 64], we extract a 5-core dense subset covering years 2018-2021 for Yelp datasets, including only users and items with more than 5 interactions. Its users potentially experienced drastic changes in the outlier environment [51]. To show such drift, we divide these datasets evenly into 10 stages in chronological order and measure the popularity drift between successive stages [71]. We firstly collect the relative popularity of items in each stage $[p^i_1, p^i_2, \cdots, p^i_t]$. Then we apply Jensen-Shannon Divergence (JSD) [19, 71] to measure the similarity between two distributions in different stages. The trends in Fig. 2 show Yelp contains a larger scale of drift while Diginetica and Kuai can be seen as normal situations. The statistics of these datasets are summarized in Table 1. We partitioned the datasets into three sets - training, validation,

and testing - in chronological order using Leave-one-out for each user, applied to all baselines.

*4.1.2 Baselines* Nine baselines from four main categories are selected in our experiments, including: *Conventional*: MF [32]; *Debiased*: MF_REL [46], DICE [72], PDA [71]; *Sequential*: GRU4Rec [27], SASRec [29]; *Graph based*: LESSR [12], GCEGNN [60], NISER [23].

*4.1.3 Metrics* $n$DCG is applied to measure the accuracy of recommendations; and Mean Intra User Distance (MIUD) [8], Tail Percentage (Tail) [37] are used to measure *intra-user* and *inter-user* diversity. We collect top-$k$ = 2 and 10 recommendation results to calculate these metrics.

### 4.2 Performance Comparison

As shown in Table 2, the proposed TCHN outperforms all baselines in both accuracy and diversity from multiple perspectives.

*4.2.1 Performance in Normal Situations* The results on Diginetica and Kuai datasets demonstrate the advantage of sequential and GNN-based models over debiased models. However, the proposed TCHN achieves great improvement when compared with baselines. This is because TCHN disentangles users' interest and conformity signals from their historical interactions, allowing them to play distinct roles in prediction based on their unique characteristics. As a result, TCHN can recommend niche items that align with users' interests rather than solely popular ones. It is particularly evident in short video datasets (Kuai), as social attributes are prevalent on these platforms, making users more susceptible to influence and potentially leading to the development of new interests and tastes.

*4.2.2 Performance in Unusual Situations* The Yelp (2018-2021) dataset effectively captures the population-scale shift in user behavior during the pandemic [51]. Conventional debiased methods falter in generating personalized recommendations in such context, due to outdated user preferences [39]. Specifically, all baseline methods encounter difficulties in offering diverse recommendations when users exhibit homogeneous behavioral patterns. However, our model sustains optimal diversity without compromising accuracy. It adeptly capitalizes on users' conformity behavior and aligns their previous interests to deliver effective personalized recommendations.

### 4.3 Ablation Studies

Given the four components demonstrated in Section 3, we conduct several ablation studies by insulating different components.

*4.3.1 Impact Temporal Relative Position Encoding* Inspired by the Hawkes process, we introduce the temporal relative position (TRP) encoding method, which enhances the temporal awareness capabilities of traditional attention mechanisms. To validate the effectiveness of this approach, we substitute our position encoding with other commonly employed techniques such as order-position (OP) encoding in [29] and reversed position (RP) encoding in [60]. Table 3 illustrates the performance of various methods across three datasets. Clearly, the temporal relative position encoding surpasses other comparative settings, as it adeptly captures the temporal dependencies within interaction sequences.

---

[3]https://competitions.codalab.org/competitions/11161
[4]https://www.kaggle.com/datasets/yelp-dataset/yelp-dataset

**Table 2: Performance on Three Datasets: Best results are highlighted in bold, while the second-best results are underlined.**

| | Normal Situations | | | | | | | | Unusual Situations (e.g., Pandemic) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | Diginetica | | | | Kuai | | | | Yelp | | | |
| | nDCG@2 | nDCG@10 | MIUD@10 | Tail@10 | nDCG@2 | nDCG@10 | MIUD@10 | Tail@10 | nDCG@2 | nDCG@10 | MIUD@10 | Tail@10 |
| MF | 0.0574 | 0.1730 | 0.583 | 0.0051 | 0.1856 | 0.2102 | 0.660 | 0.0059 | 0.0041 | 0.0076 | 0.601 | 0.0045 |
| MF_REL | 0.0659 | 0.1785 | 0.578 | 0.0058 | 0.1904 | 0.2212 | 0.656 | 0.0065 | 0.0045 | 0.0085 | 0.703 | 0.0049 |
| DICE | 0.0984 | 0.1989 | 0.588 | 0.0089 | 0.2168 | 0.2503 | 0.655 | 0.0099 | 0.0062 | 0.0101 | 0.703 | 0.0054 |
| PDA | 0.1172 | 0.2004 | 0.566 | 0.0095 | 0.2203 | 0.2565 | 0.623 | 0.0101 | 0.0058 | 0.0095 | 0.699 | 0.0056 |
| GRU4Rec | 0.1271 | 0.2142 | 0.604 | 0.0154 | 0.2545 | 0.2850 | 0.697 | 0.0250 | 0.0134 | 0.0278 | 0.693 | 0.0171 |
| SASRec | 0.1406 | 0.2312 | 0.604 | 0.0166 | 0.2587 | 0.2878 | 0.693 | 0.0280 | 0.0158 | 0.0332 | 0.697 | 0.0174 |
| LESSR | 0.1391 | 0.2297 | 0.603 | 0.0143 | 0.2618 | 0.2907 | 0.696 | 0.0276 | 0.0154 | 0.0303 | 0.688 | 0.0079 |
| NISER | 0.1506 | 0.2432 | 0.603 | 0.0131 | 0.2576 | 0.2862 | 0.696 | 0.0274 | 0.0153 | 0.0314 | 0.688 | 0.0077 |
| GCEGNN | 0.1555 | 0.2489 | 0.603 | 0.0178 | 0.2569 | 0.2891 | 0.696 | 0.0268 | 0.0177 | 0.0350 | 0.688 | 0.0174 |
| THN-si | 0.1546 | 0.2452 | 0.603 | 0.0171 | 0.2646 | 0.2902 | 0.696 | 0.0424 | 0.0183 | 0.0354 | 0.658 | 0.0077 |
| THN-sn | 0.1323 | 0.2132 | 0.611 | 0.0146 | 0.2577 | 0.2871 | 0.706 | 0.0373 | 0.0144 | 0.0328 | 0.698 | 0.0177 |
| TCHN | **0.1615** | **0.2510** | **0.624** | **0.0196** | **0.2686** | **0.2989** | **0.703** | **0.0476** | **0.0197** | **0.0366** | **0.724** | **0.0182** |
| *p*-value | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.01 | <0.001 | <0.001 |

**Table 3: Comparison of position encoding methods**

| | Diginetica | Kuai | Yelp |
|---|---|---|---|
| | nDCG@10 | nDCG@10 | nDCG@10 |
| TRP | **0.2510** | **0.2989** | **0.0366** |
| OP | 0.2340 | 0.2901 | 0.0345 |
| RP | 0.2485 | 0.2920 | 0.0350 |

*4.3.2 Conformity Modeling* This section demonstrates the effect of *informational conformity* and *normative conformity* components. We eliminate the normative conformity layer, denoted as "THN-si", and the informational conformity layer, denoted as "THN-sn" respectively. As shown in Table 2, "THN-si" can still outperform other baselines w.r.t accuracy and diversity on Diginetica and Kuai datasets, which reveals the feasibility and effectiveness of integrating the informational conformity model into TCHN. However, it is unable to detect irrational user behaviors in interactions due to similar limitations as conventional RS models. As a result, it loses too much diversity score in the Yelp dataset, potentially misinterpreting normative conformity behavior as users' genuine interest. On the other hand, "THN-sn" fails to generate comparable recommendations as baselines due to its failure to acknowledge the positive impact of like-minded neighbors on the user in question. In "THN-sn", users cannot get nourishment and develop new interests from those who they think are knowledgeable predecessors. In summary, TCHN can significantly improve upon incomplete models, particularly in short-video data and unusual situations.

## 4.4 Studies on Different User Groups

As aforementioned, we define two groups of users: "surfers" and "deep divers" based on their perception of conformity and taste for popular items. As shown in Fig. 3, we evaluate our model on different user groups to confirm that the recommended items can benefit all users with various perceptions of conformity and taste for popular items. Here, we divided users into ten groups according to the normalized average popularity of the items they interacted with. For each group, we measure the *n*DCG@10 of the recommendations. We can see that TCHN achieves consistent performance across all groups of users, while other baselines clearly favor "surfers" who like popular items.
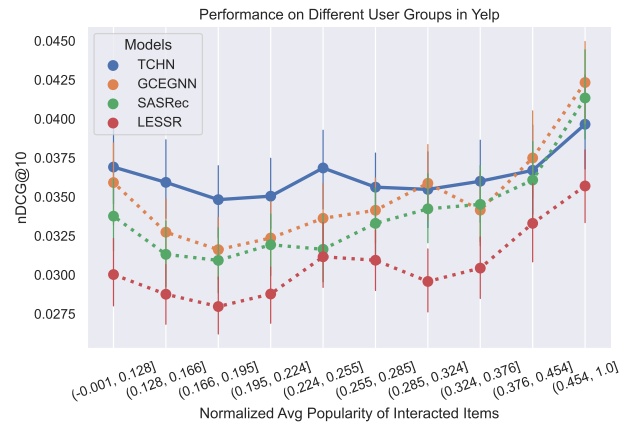


**Figure 3: Performance on Different User Groups in Yelp**

## 5 CONCLUSION

Aiming at the problem that conventional RSs potentially ignore the positive effect of users' conformity behavior, we propose a TCHN for disentangling user interest and conformity based on attentional Hawkes process networks. This paper investigates how to model the user interest and the conformity evolution at the individual level to maximize their effect for recommendation purposes, and the experiment results show that modeling in this way leads to better recommendations in terms of accuracy and diversity.

## 6 ACKNOWLEDGMENTS

## REFERENCES

[1] Clayton P. Alderfer. 1969. An Empirical Test of a New Theory of Human Needs. *Organizational Behavior and Human Performance* 4, 2 (May 1969), 142–175. https://doi.org/10.1016/0030-5073(69)90004-X
[2] Elliot Aronson. 1969. The theory of cognitive dissonance: A current perspective. In *Advances in experimental social psychology*. Vol. 4. Elsevier, 1–34.

[3] Ting Bai, Lixin Zou, Wayne Xin Zhao, Pan Du, Weidong Liu, Jian-Yun Nie, and Ji-Rong Wen. 2019. CTrec: A long-short demands evolution model for continuous-time recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 675–684.

[4] Russell W Belk. 1974. An exploratory assessment of situational effects in buyer behavior. *Journal of marketing research* 11, 2 (1974), 156–163.

[5] Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy* 100, 5 (1992), 992–1026.

[6] Stephen Bonner and Flavian Vasile. 2018. Causal embeddings for recommendation. In *Proceedings of the 12th ACM conference on recommender systems*. 104–112.

[7] Rocío Cañamares and Pablo Castells. 2018. Should I follow the crowd? A probabilistic analysis of the effectiveness of popularity in recommender systems. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 415–424.

[8] Pablo Castells, Neil Hurley, and Saul Vargas. 2022. Novelty and diversity in recommender systems. In *Recommender systems handbook*. Springer, 603–646.

[9] Jianxin Chang, Chen Gao, Yu Zheng, Yiqun Hui, Yanan Niu, Yang Song, Depeng Jin, and Yong Li. 2021. Sequential recommendation with graph neural networks. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*. 378–387.

[10] Jiawei Chen, Hande Dong, Xiang Wang, Fuli Feng, Meng Wang, and Xiangnan He. 2020. Bias and debias in recommender system: A survey and future directions. *ACM Transactions on Information Systems* (2020).

[11] Qiwei Chen, Huan Zhao, Wei Li, Pipei Huang, and Wenwu Ou. 2019. Behavior sequence transformer for e-commerce recommendation in alibaba. In *Proceedings of the 1st International Workshop on Deep Learning Practice for High-Dimensional Sparse Data*. 1–4.

[12] Tianwen Chen and Raymond Chi-Wing Wong. 2020. Handling information loss of graph neural networks for session-based recommendation. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1172–1180.

[13] Robert B Cialdini and Noah J Goldstein. 2004. Social influence: Compliance and conformity. *Annu. Rev. Psychol.* 55 (2004), 591–621.

[14] Nicolas Claidière and Andrew Whiten. 2012. Integrating the study of conformity and culture in humans and nonhuman animals. *Psychological bulletin* 138, 1 (2012), 126.

[15] Daryl J Daley and David Vere-Jones. 2007. *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Science & Business Media.

[16] Morton Deutsch and Harold B Gerard. 1955. A study of normative and informational social influences upon individual judgment. *The journal of abnormal and social psychology* 51, 3 (1955), 629.

[17] Tim Donkers, Benedikt Loepp, and Jürgen Ziegler. 2017. Sequential user-based recurrent neural network recommendations. In *Proceedings of the eleventh ACM conference on recommender systems*. 152–160.

[18] Nan Du, Hanjun Dai, Rakshit Trivedi, Utkarsh Upadhyay, Manuel Gomez-Rodriguez, and Le Song. 2016. Recurrent marked temporal point processes: Embedding event history to vector. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 1555–1564.

[19] Dominik Maria Endres and Johannes E Schindelin. 2003. A new metric for probability distributions. *IEEE Transactions on Information theory* 49, 7 (2003), 1858–1860.

[20] Andrew J Flanagin and Miriam J Metzger. 2013. Trusting expert-versus user-generated ratings online: The role of information volume, valence, and consumer characteristics. *Computers in Human Behavior* 29, 4 (2013), 1626–1634.

[21] Chongming Gao, Shijun Li, Wenqiang Lei, Jiawei Chen, Biao Li, Peng Jiang, Xiangnan He, Jiaxin Mao, and Tat-Seng Chua. 2022. KuaiRec: A Fully-observed Dataset and Insights for Evaluating Recommender Systems. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management* (Atlanta, GA, USA) *(CIKM '22)*. 11 pages. https://doi.org/10.1145/3511808.3557220

[22] Jacob K Goeree, Thomas R Palfrey, Brian W Rogers, and Richard D McKelvey. 2007. Self-correcting information cascades. *The Review of Economic Studies* 74, 3 (2007), 733–762.

[23] Priyanka Gupta, Diksha Garg, Pankaj Malhotra, Lovekesh Vig, and Gautam Shroff. 2019. NISER: Normalized item and session representations to handle popularity bias. *arXiv preprint arXiv:1909.04276* (2019).

[24] Alan G Hawkes. 1971. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58, 1 (1971), 83–90.

[25] Yue He, Zimu Wang, Peng Cui, Hao Zou, Yafeng Zhang, Qiang Cui, and Yong Jiang. 2022. CausPref: Causal Preference Learning for Out-of-Distribution Recommendation. In *Proceedings of the ACM Web Conference 2022* (Virtual Event, Lyon, France) *(WWW '22)*. Association for Computing Machinery, New York, NY, USA, 410–421. https://doi.org/10.1145/3485447.3511969

[26] Peter Hedström and Petri Ylikoski. 2010. Causal mechanisms in the social sciences. *Annual review of sociology* 36 (2010), 49–67.

[27] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2015. Session-based recommendations with recurrent neural networks. *arXiv*

[preprint arXiv:1511.06939](2015).

[28] Kosuke Imai, Luke Keele, Dustin Tingley, and Teppei Yamamoto. 2011. Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review* 105, 4 (2011), 765–789.

[29] Wang-Cheng Kang and Julian McAuley. 2018. Self-attentive sequential recommendation. In *2018 IEEE international conference on data mining (ICDM)*. IEEE, 197–206.

[30] Harold H Kelley and Martin M Shapiro. 1954. An experiment on conformity to group norms where conformity is detrimental to group achievement. *American Sociological Review* 19, 6 (1954), 667–677.

[31] Yehuda Koren. 2009. Collaborative filtering with temporal dynamics. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. 447–456.

[32] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 42, 8 (2009), 30–37.

[33] Sanjay Krishnan, Jay Patel, Michael J Franklin, and Ken Goldberg. 2014. A methodology for learning, analyzing, and mitigating social influence bias in recommender systems. In *Proceedings of the 8th ACM Conference on Recommender systems*. 137–144.

[34] Hui Li, Hui Li, and Sourav S Bhowmick. 2020. Chassis: Conformity meets online information diffusion. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*. 1829–1840.

[35] Jiacheng Li, Yujie Wang, and Julian McAuley. 2020. Time interval aware self-attention for sequential recommendation. In *Proceedings of the 13th international conference on web search and data mining*. 322–330.

[36] Yujia Li, Daniel Tarlow, Marc Brockschmidt, and Richard Zemel. 2015. Gated graph sequence neural networks. *arXiv preprint arXiv:1511.05493* (2015).

[37] Siyi Liu and Yujia Zheng. 2020. Long-tail session-based recommendation. In *Proceedings of the 14th ACM Conference on Recommender Systems*. 509–514.

[38] Chenglong Ma, Yongli Ren, Pablo Castells, and Mark Sanderson. 2022. Evaluation of Herd Behavior Caused by Population-Scale Concept Drift in Collaborative Filtering. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Madrid, Spain) *(SIGIR '22)*. Association for Computing Machinery, New York, NY, USA, 1984–1989. https://doi.org/10.1145/3477495.3531792

[39] Chenglong Ma, Yongli Ren, Pablo Castells, and Mark Sanderson. 2022. NEST: Simulating Pandemic-like Events for Collaborative Filtering by Modeling User Needs Evolution. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management* (Atlanta, GA, USA) *(CIKM '22)*. Association for Computing Machinery, New York, NY, USA, 1430–1440. https://doi.org/10.1145/3511808.3557407

[40] A. H. Maslow. 1943. A Theory of Human Motivation. *Psychological Review* 50, 4 (1943), 370–396. https://doi.org/10.1037/h0054346

[41] Abraham H. Maslow. 1970. New Introduction: Religions, Values, and Peak-Experiences. *Journal of Transpersonal Psychology* 2, 2 (1970), 83–90.

[42] Hongyuan Mei and Jason M Eisner. 2017. The neural hawkes process: A neurally self-modulating multivariate point process. *Advances in neural information processing systems* 30 (2017).

[43] J Payne, JR Bettman, and EJ Johnson. 1991. Consumer decision making. *Handbook of consumer behaviour* (1991), 50–84.

[44] Andrew K Przybylski, Kou Murayama, Cody R DeHaan, and Valerie Gladwell. 2013. Motivational, emotional, and behavioral correlates of fear of missing out. *Computers in human behavior* 29, 4 (2013), 1841–1848.

[45] Izhak Rubin. 1972. Regular point processes and their detection. *IEEE Transactions on Information Theory* 18, 5 (1972), 547–557.

[46] Yuta Saito, Suguru Yaginuma, Yuta Nishino, Hayato Sakata, and Kazuhide Nakata. 2020. Unbiased recommender learning from missing-not-at-random implicit feedback. In *Proceedings of the 13th International Conference on Web Search and Data Mining*. 501–509.

[47] Markus Schedl. 2019. Deep learning in music recommendation systems. *Frontiers in Applied Mathematics and Statistics* (2019), 44.

[48] Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. 2016. Recommendations as treatments: Debiasing learning and evaluation. In *international conference on machine learning*. PMLR, 1670–1679.

[49] Guy Shani and Asela Gunawardana. 2011. *Evaluating Recommendation Systems*. Springer US, Boston, MA, 257–297. https://doi.org/10.1007/978-0-387-85820-3_8

[50] Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. 2018. Self-attention with relative position representations. *arXiv preprint arXiv:1803.02155* (2018).

[51] Jina Suh, Eric Horvitz, Ryen W. White, and Tim Althoff. 2021. Population-Scale Study of Human Needs During the COVID-19 Pandemic: Analysis and Implications *(WSDM '21)*. Association for Computing Machinery, New York, NY, USA, 4–12. https://doi.org/10.1145/3437963.3441788

[52] Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. 2019. BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. In *Proceedings of the 28th ACM international conference on information and knowledge management*. 1441–1450.

[53] Henri Tajfel, John C Turner, William G Austin, and Stephen Worchel. 1979. An integrative theory of intergroup conflict. *Organizational identity: A reader* 56, 65 (1979), 9780203505984–16.

[54] Yong Kiam Tan, Xinxing Xu, and Yong Liu. 2016. Improved recurrent neural networks for session-based recommendations. In *Proceedings of the 1st workshop on deep learning for recommender systems*. 17–22.

[55] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[56] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. *International Conference on Learning Representations* (2018). https://openreview.net/forum?id=rJXMpikCZ

[57] Chenyang Wang, Min Zhang, Weizhi Ma, Yiqun Liu, and Shaoping Ma. 2019. Modeling item-specific temporal dynamics of repeat consumption for recommender systems. In *The World Wide Web Conference*. 1977–1987.

[58] Dongjing Wang, Xin Zhang, Zhengzhe Xiang, Dongjin Yu, Guandong Xu, and Shuiguang Deng. 2021. Sequential recommendation based on multivariate Hawkes process embedding with attention. *IEEE transactions on cybernetics* 52, 11 (2021), 11893–11905.

[59] Jianling Wang, Kaize Ding, Ziwei Zhu, and James Caverlee. 2021. Session-based recommendation with hypergraph attention networks. In *Proceedings of the 2021 SIAM International Conference on Data Mining (SDM)*. SIAM, 82–90.

[60] Ziyang Wang, Wei Wei, Gao Cong, Xiao-Li Li, Xian-Ling Mao, and Minghui Qiu. 2020. Global context enhanced graph neural networks for session-based recommendation. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*. 169–178.

[61] Tianxin Wei, Fuli Feng, Jiawei Chen, Ziwei Wu, Jinfeng Yi, and Xiangnan He. 2021. Model-agnostic counterfactual reasoning for eliminating popularity bias in recommender system. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 1791–1800.

[62] Le Wu, Junwei Li, Peijie Sun, Richang Hong, Yong Ge, and Meng Wang. 2020. Diffnet++: A neural influence and interest diffusion network for social recommendation. *IEEE Transactions on Knowledge and Data Engineering* 34, 10 (2020), 4753–4766.

[63] Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. 2019. Session-based recommendation with graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 346–353.

[64] Chengfeng Xu, Pengpeng Zhao, Yanchi Liu, Victor S Sheng, Jiajie Xu, Fuzhen Zhuang, Junhua Fang, and Xiaofang Zhou. 2019. Graph contextualized self-attention network for session-based recommendation.. In *IJCAI*, Vol. 19. 3940–3946.

[65] Da Xu, Chuanwei Ruan, Evren Korpeoglu, Sushant Kumar, and Kannan Achan. 2020. Inductive representation learning on temporal graphs. *arXiv preprint arXiv:2002.07962* (2020).

[66] Junchi Yan. 2019. Recent advance in temporal point process: from machine learning perspective. *SJTU Technical Report* (2019).

[67] Haochao Ying, Fuzhen Zhuang, Fuzheng Zhang, Yanchi Liu, Guandong Xu, Xing Xie, Hui Xiong, and Jian Wu. 2018. Sequential recommender system based on hierarchical attention network. In *IJCAI International Joint Conference on Artificial Intelligence*.

[68] Kum Fai Yuen, Xueqin Wang, Fei Ma, and Kevin X Li. 2020. The psychological causes of panic buying following a health crisis. *International journal of environmental research and public health* 17, 10 (2020), 3513.

[69] Fue Zeng, Ran Tao, Yanwu Yang, and Tingting Xie. 2017. How social communications influence advertising perception and response in online communities? *Frontiers in psychology* 8 (2017), 1349.

[70] Qiang Zhang, Aldo Lipani, Omer Kirnap, and Emine Yilmaz. 2020. Self-attentive Hawkes process. In *International conference on machine learning*. PMLR, 11183–11193.

[71] Yang Zhang, Fuli Feng, Xiangnan He, Tianxin Wei, Chonggang Song, Guohui Ling, and Yongdong Zhang. 2021. Causal Intervention for Leveraging Popularity Bias in Recommendation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, Canada) *(SIGIR '21)*. Association for Computing Machinery, New York, NY, USA, 11–20. https://doi.org/10.1145/3404835.3462875

[72] Yu Zheng, Chen Gao, Xiang Li, Xiangnan He, Yong Li, and Depeng Jin. 2021. Disentangling User Interest and Conformity for Recommendation with Causal Embedding. In *Proceedings of the Web Conference 2021* (Ljubljana, Slovenia) *(WWW '21)*. Association for Computing Machinery, New York, NY, USA, 2980–2991. https://doi.org/10.1145/3442381.3449788

[73] Guorui Zhou, Na Mou, Ying Fan, Qi Pi, Weijie Bian, Chang Zhou, Xiaoqiang Zhu, and Kun Gai. 2019. Deep interest evolution network for click-through rate prediction. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 5941–5948.

[74] Simiao Zuo, Haoming Jiang, Zichong Li, Tuo Zhao, and Hongyuan Zha. 2020. Transformer hawkes process. In *International conference on machine learning*. PMLR, 11692–11702.

## A APPENDIX: PROOF OF POSITIVE DEFINITENESS OF THE KERNEL FUNCTION

In this appendix, we provide a proof that the kernel function $\kappa = \theta_z \ln \left( \beta_{a,b} \left( |t_b - t_a| \right) + 1 \right)$ is positive definite, given $\theta > 0$ and $\beta > 0$, thus satisfying the assumption of Bochner's theorem.

First, let us recall the definition of Bochner's theorem:

*Definition A.1 (Bochner's Theorem).* A complex-valued function $f$ on a locally compact abelian group $G$ is positive definite if and only if $f$ is the Fourier transform of a positive measure on the dual group $\hat{G}$.

Proof. Let $N \in \mathbb{N}$ and $t_1, t_2, \ldots, t_N \in \mathbb{R}^+$ be arbitrary. We must show that the matrix $[\kappa(t_i - t_j)]_{i,j=1}^N$ is positive semi-definite. This means for any vector $x = (x_1, x_2, \ldots, x_N) \in \mathbb{R}^N$, we have

$$\sum_{i=1}^N \sum_{j=1}^N x_i \overline{x_j} \kappa(t_i - t_j) \geq 0.$$

Substituting the expression for $\kappa$, we get

$$\sum_{i=1}^N \sum_{j=1}^N x_i \overline{x_j} \theta_z \ln \left( \beta_{a,b} \left( |t_i - t_j| \right) + 1 \right) \geq 0.$$

Since $\theta_z > 0$ and $\ln(x + 1)$ is increasing for $x \geq 0$, it suffices to show that $\sum_{i=1}^N \sum_{j=1}^N x_i \overline{x_j} \beta_{a,b} \left( |t_i - t_j| \right) \geq 0$.

This follows from the fact that the function $|t_i - t_j|$ is non-negative, and the product of non-negative numbers is non-negative. Hence, the kernel function $\kappa = \theta_z \ln \left( \beta_{a,b} \left( |t_b - t_a| \right) + 1 \right)$ is positive definite. □

This confirms that our kernel function satisfies the assumptions of Bochner's theorem, and can therefore be used in the context of harmonic analysis and related mathematical fields.