Where To Next? A Dynamic Model of User Preferences

Francesco Sanna Passino*
Imperial College London
fs4515@ic.ac.uk

Lucas Maystre
Spotify
lucasm@spotify.com

Dmitrii Moor
Spotify
dmitriim@spotify.com

Ashton Anderson†
University of Toronto
ashton@cs.toronto.edu

Mounia Lalmas
Spotify
mounia@acm.org

ABSTRACT

We consider the problem of predicting users' preferences on online platforms. We build on recent findings suggesting that users' preferences change over time, and that helping users expand their horizons is important in ensuring that they stay engaged. Most existing models of user preferences attempt to capture *simultaneous* preferences: "Users who like A tend to like B as well". In this paper, we argue that these models fail to anticipate changing preferences. To overcome this issue, we seek to understand the structure that underlies the *evolution* of user preferences. To this end, we propose the Preference Transition Model (PTM), a dynamic model for user preferences towards classes of items. The model enables the estimation of transition probabilities between classes of items over time, which can be used to estimate how users' tastes are expected to evolve based on their past history. We test our model's predictive performance on a number of different prediction tasks on data from three different domains: music streaming, restaurant recommendations and movie recommendations, and find that it outperforms competing approaches. We then focus on a music application, and inspect the structure learned by our model. We find that the PTM uncovers remarkable regularities in users' preference trajectories over time. We believe that these findings could inform a new generation of dynamic, diversity-enhancing recommender systems.

KEYWORDS

Recommender systems, time series, diversity, user modelling.

ACM Reference Format:
Francesco Sanna Passino, Lucas Maystre, Dmitrii Moor, Ashton Anderson, and Mounia Lalmas. 2021. Where To Next? A Dynamic Model of User Preferences. In *Proceedings of the Web Conference 2021 (WWW '21), April 19–23, 2021, Ljubljana, Slovenia.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3442381.3450028

*This work was completed as part of an internship at Spotify.
†This work was completed while the author was working at Spotify.

WWW '21, April 19–23, 2021, Ljubljana, Slovenia

ACM ISBN 978-1-4503-8312-7/21/04.
https://doi.org/10.1145/3442381.3450028

1 INTRODUCTION

Online platforms have transformed the way users access information, audio and video content, knowledge repositories, and much more. Over three decades of research and practice have demonstrated that *a*) learning users' preferences, and *b*) personalizing users' experience to match these preferences is immensely valuable to increase engagement and satisfaction. To this end, recommender systems have emerged as essential building blocks [3]. They help users find their way through large collections of items and assist them in discovering new content. They typically build on user preference models that exploit correlations across users' preferences. As an example within the music domain, if a user likes The Beatles, that user might also like Simon & Garfunkel, because other users who listen to the former also listen to the latter.

The vast majority of these preference models take a static viewpoint and try to capture *simultaneous* preferences. In other words, they seek to answer the following question: How well does preference for item A predict preference for item B? While these models are effective at finding similar items, they can confine users in so-called *filter bubbles* [30, 31]. In a sense, they successfully *exploit* narrow preferences but may not help users *explore* completely new regions in the space of items. From a user's point of view, this can lead to the impression that recommender systems tend to show "more of the same" [27], and that they do not help expand their horizons.

As online platforms mature, our understanding of users' preferences and behaviors is becoming richer and more nuanced. There is evidence suggesting that traditional approaches, based on matching users with items similar to those they have liked in the past, fail to capture important aspects of long-term engagement. In particular, the following two findings stand out.

(1) Longitudinal traces reveal that the journey of users on online platform is inherently *dynamic*, and that their preferences are fluid [23]. Take a music streaming service, for example: new artists and tracks become available over time, trends evolve quickly and users' preferences with them.
(2) Consuming *diverse* content is linked to users' engagement and long-term satisfaction [10, 15, 20, 35]. In the context of a music streaming service, for example, recent research shows that consumption diversity is highly correlated with user retention and conversion to a paid subscription [4].

Previous research has attempted to address these aspects without fundamentally changing the underlying preference model. To account for preference drift over time, dynamic variants of well-known preference models have been proposed [9, 23]. These models alleviate the staleness problem, but they do not capture regularities within the preference trajectories themselves. In other words, they recognize that users' preferences *today* might be different from *a year ago*, but they cannot be used to predict how these preferences will look like *one year from now*. To address the lack of diversity in recommendations, some proposals suggest adding constraints such that sets of recommended items cover distinct regions of the item space [1], or framing recommendation as a multi-objective problem [39]. These methods may succeed in exposing users to diverse items, but they again fail to capture *how* users discover new items. Arguably, the latter is crucial to ensure that novel recommended items remain relevant.

In this work, we overcome these issues by modeling preference trajectories *explicitly*. We assume that we have access to timestamped consumption traces for a set of users, and we ask the following question.

Question. *How well does preference for item A at time t predict preference for item B at time t + 1?*

In contrast to most existing preference models, this question examines the structure of preferences' temporal dynamics. It suggests that users' changing preferences exhibit some regularities, and that we might be able to learn these regularities in a data-driven way.

To this end, we propose the Preference Transition Model (PTM), a dynamic model of preferences that predicts how users move in the space of items. At its core is a transition matrix that captures the propensity of interacting with a given item at time $t + 1$ given all interactions up to time t; this transition matrix provides a rich representation of the underlying preference dynamics and is highly interpretable. By design, our model anticipates changes in users' preferences, and accompanies them through their journey. Furthermore, by taking advantage of the transition structure our model uncovers, we can gradually bring users towards unexplored regions of item space, thus increasing diversity in a meaningful way.

We validate our model's performance on several tasks using four datasets covering three different application domains. We show that it predicts users' future preferences more accurately than competing approaches. Beyond accurate predictions, we find that our model can identify promising pathways between potentially distant regions in the space of items. We illustrate this point in Figure 1. We fit our model on a dataset of music streams and, based on the learned transition matrix, we compute the most likely trajectory between any two musical genres (details given in Section 5.2.1). These trajectories reveal valuable insights that could be turned into principled, diversity-enhancing recommendations.

Organization of the Paper. After discussing the relevant literature in Section 2, we introduce the PTM in Section 3. We evaluate its performance on several datasets in Section 4 and compare it to alternative approaches. In Section 5, we dive deeper into an application to music preferences, and we demonstrates that the model discovers meaningful structure in the music space. Finally, we conclude in Section 6.

2 RELATED WORK

This work is related to several different areas, each of which we discuss here, emphasising connections and differences.

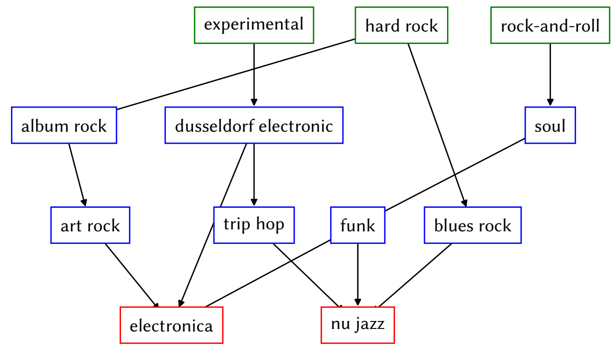

Figure 1: Given a dataset of music consumption traces, the PTM automatically finds the most likely trajectories from source genres (in green) to destination genres (in red).

Dynamic Collaborative Filtering. Our contribution can be broadly classified within the work on dynamic collaborative filtering in the context of matrices of counts. Traditionally, models for collaborative filtering give users and items a low-dimensional latent feature representation. One popular approach for modelling large sparse matrices of counts is Poisson Matrix Factorisation (PMF, [16]), which assigns a non-negative latent feature representation to users and items. Matrix factorization based on the minimisation of the Frobenius norm has also been used successfully [36, 40]. Furthermore, many authors have recognised the importance of incorporating time-evolving features within models for collaborative filtering. Koren [23] proposed the popular *timeSVD++* method, a dynamic matrix factorisation technique. Also, PMF has been extended to allow for dynamically evolving latent features in [9]. Similarly, Schein et al. [33] use a Bayesian tensor approach to account for temporal variations in the latent representation. Tensor factorisation methods are also successfully developed in [21, 38]. Acharya et al. [2] assume a Gamma Process structure for temporal changes in the latent features, but their inferential procedure is not easily scalable to large numbers of users or items. Similarly, Li et al. [25] propose Dirichlet-distributed latent representations with a Markov prior structure. Gaussian-based models with Kalman filter updates for the latent features have also been successfully developed by Gultekin and Paisley [17]. The main difference between our PTM and existing methodologies is that in this work the distribution of user activity across classes of items and the total number of interactions in each time period are modelled *separately*, using a two-stage Poisson-Multinomial hierarchical distribution. This approach has the advantage to suitably account for heterogeneity in user activity over time.

Sequential Recommender Systems. Some authors have addressed variants of the dynamic recommendation problem by using sequential models, in particular recurrent neural networks (RNNs). Wu et al. [37] introduce recurrent recommender networks, and Beutel et al. [7] discuss how to model context in RNN-based recommender systems. RNNs are also commonly used in the context of session-based recommender systems such as GRU4Rec [19]. In this work, we are mostly interested in long-term shifts in user preferences as opposed to short-term sessions, and we focus on obtaining interpretable insights into how user preferences change over time. These differentiate our PTM framework from deep-learning based sequential models proposed in the literature.

Diversity in Recommender Systems. One of the possible applications of the model proposed in this paper is to provide more diverse recommendations to users, as well as promoting new classes of items for consumption. Traditionally, diversity is included in recommender systems in a multi-objective framework. In this work, a different approach is taken, and the diversity stems from recommendations of new classes of items to explore. An extensive review of the traditional approaches for promoting diversity within recommender systems can be found in [24] and references therein. Particularly relevant for our application to music streaming services is [20], which establishes a connection between engagement and diversity on a streaming platform.

Estimation of Relationships Between Multivariate Processes. As described in Section 1, the PTM model proposed in this work is aimed at estimating relationships between classes of items within recommender systems. Our approach closely resembles techniques used in statistics and causal inference for estimating relationships between multivariate processes. In particular, vector autoregressive (VAR) processes are traditionally used to establish causal interactions between time series [5, 14]. In Section 4.3, we compare our PTM model to VARs. Similarly, Hawkes processes have also been used for the same purpose on point process data [26]. A more general class of models are self-exciting point processes (SEPPs, [28]), which could also be used for estimation of the interactions. Both VAR and Hawkes processes have good theoretical properties for estimation of a network structure between different processes, both in terms of Granger causality and directed information [13, 14].

Qualitative Interpretation of Interaction Graphs. Finally, part of this work could be related to techniques for extracting qualitatively meaningful network relationships between entities. In particular, within a music streaming application, the PTM model is used to discover pathways between genres and how the users explored the music space. The work of Shahaf and Guestrin [34], which describes an algorithm for finding meaningful connections between news articles, is particularly relevant in this context, and provided ideas for the analysis in Section 5.

Others. Our work could also be linked to that of Dunlavy et al. [12], where, in the context of a link prediction task, a weighted combination of previous observations is considered and subsequently fed into a matrix factorization framework. Finally, the model proposed in this article could be also related to the work of McAuley and Leskovec [29], where the dynamic evolution of user preferences is related to their *expertise*, such that user tastes are expected to change based on the number and type of items consumed in the past. In this work, the dynamics is induced by an exponentially weighted moving average distribution, which gives more weight to recent observations.

3 PREFERENCE TRANSITION MODEL

In this section, we develop the Preference Transition Model (PTM). We first introduce some notation and formally define our prediction problem. Then, we proceed to give a detailed description of the model and related estimation and prediction methods.

3.1 Notation & Problem Statement

We consider a set of M users interacting with items over time. We assume that each item belongs to one of N classes, and we divide the observation period into T time windows. The exact definition of a *class* is domain-dependent. As a running example, we consider a music streaming application where users listen to tracks of different musical genres, and we let the item classes be the set of genres. We let $n_{ij}^t \in \mathbb{N} \cup \{0\}$ be the number of times the user i interacted with an item in class j within the time window t. For example, in the music streaming domain, n_{ij}^t represents the number of tracks of genre j streamed by user i in the time window t. We define the vector of streams for user i at time t as

$$\boldsymbol{n}_i^t = \begin{bmatrix} n_{i1}^t & \cdots & n_{iN}^t \end{bmatrix}. \tag{1}$$

Furthermore, the *activity* of user i in the t-th time window is the total number of interactions within the time period:

$$\xi_i^t = \sum_{j=1}^{N} n_{ij}^t. \tag{2}$$

We assume that $\xi_i^t > 0$ for all users $i = 1, \dots, M$ and all time windows $t = 1, \dots, T$. Combining n_{ij}^t and ξ_i^t gives the *class distribution* at time t:

$$\boldsymbol{\pi}_i^t = \begin{bmatrix} \pi_{i1}^t & \cdots & \pi_{iN}^t \end{bmatrix} = \boldsymbol{n}_i^t / \xi_i^t. \tag{3}$$

The prediction problem we address can be formalized as follows. Given $\{\boldsymbol{n}_i^t : i, t \in \{1, \dots, M\} \times \{1, \dots, T\}\}$, predict $\boldsymbol{n}_i^{T+1}$ or $\boldsymbol{\pi}_i^{T+1}$ for every user $i \in \{1, \dots, M\}$.

3.2 Statistical Model

We now present a dynamic statistical model to study the evolution of the counts $\boldsymbol{n}_i^t$, and consequently the distributions $\boldsymbol{\pi}_i^t$, over time, establishing relationships between different classes. We call it the *Preference Transition Model (PTM)*. The PTM is aimed at jointly modelling sequences of user activities and class distributions over time.

First, we postulate that the distribution of preferences of user i across classes at time window t depends on a *combination* of the distributions at the previous time windows. In particular, we let $\gamma \in [0, 1]$ be an *exploration* parameter that measures *how quickly* the space of classes is explored on average. We assume that γ is the same for all users. Conditional on γ, an exponentially weighted moving average (EWMA, [32]) class distribution is obtained within each time window:

$$\tilde{\boldsymbol{\pi}}_i^t = (1 - \gamma)\tilde{\boldsymbol{\pi}}_i^{t-1} + \gamma \boldsymbol{\pi}_i^t, \tag{4}$$

with initial value $\tilde{\boldsymbol{\pi}}_i^1 = \boldsymbol{\pi}_i^1$. The EWMA distribution $\tilde{\boldsymbol{\pi}}_i^t$ in Equation (4) contains information about the entire history of the user trajectory in the space of items, giving more weight to recent observations, and exponentially down-weighting previous observations over time. Using $\tilde{\boldsymbol{\pi}}_i^t$, we aim to approximate $\boldsymbol{\pi}_i^{t+1}$ as closely as possible, up to the effect of the interactions between classes.

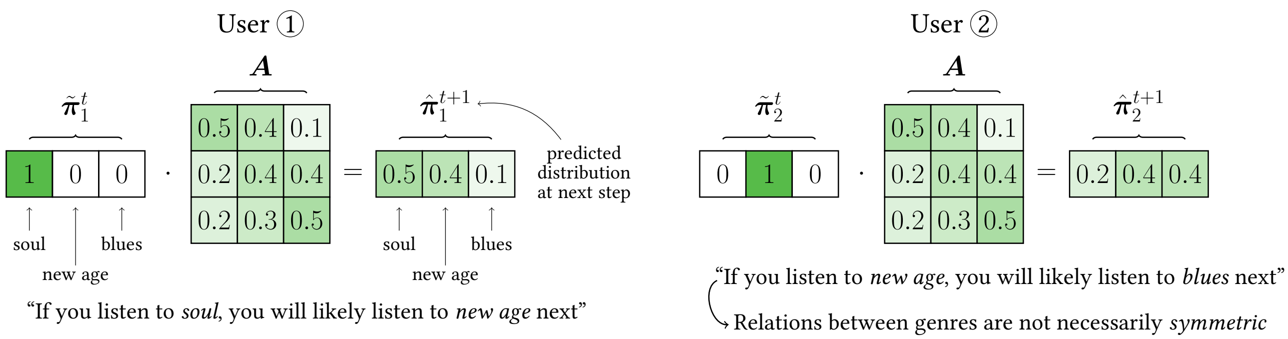

Figure 2: Visual representation of a PTM and corresponding prediction method within a music streaming application.

Second, we assume that the activity ξ_i^{t+1} of the user i at time window $t+1$ has a Poisson distribution with mean ξ_i^t, corresponding to the activity of the user in the previous time window. Furthermore, conditional on the activity at time $t+1$, the counts $\boldsymbol{n}_i^{t+1}$ are modelled using a multinomial distribution with probability given by $\tilde{\boldsymbol{\pi}}_i^t$, re-weighted by a $N\times N$ matrix of interactions between classes, encoded by a stochastic matrix $\boldsymbol{A}$. In summary, the PTM model can be written as follows:

$$\begin{aligned} \xi_i^{t+1} &\sim \text{Poisson}\left(\xi_i^t\right), \\ \boldsymbol{n}_i^{t+1} \mid \xi_i^{t+1},\ \tilde{\boldsymbol{\pi}}_i^t &\sim \text{Multinomial}\left(\xi_i^{t+1},\ \tilde{\boldsymbol{\pi}}_i^t \boldsymbol{A}\right). \end{aligned} \tag{5}$$

Here, $\boldsymbol{A} = \{a_{jk}\} \in [0,1]^{N\times N}$ is a stochastic matrix where:

(1) all the entries are non-negative: $a_{jk} \geq 0\ \forall\, j,k = 1,\dots,N$,
(2) all the rows sum up to 1: $\sum_{k=1}^N a_{jk} = 1\ \forall\, j = 1,\dots,N$.

A key feature of the PTM model in Equation (5) is that the activity and the class distribution are modelled *separately*. The formulation of the PTM in Equation (5) also closely resembles a first-order Markov chain, where the states correspond to the classes. Under this framework, assuming $\gamma = 1$, the entries of $\boldsymbol{A}$ take a simple interpretation: if users were allowed to interact only with one class at each time point, and that a given user liked class j at time t, then a_{jk} would represent the probability that the user makes a transition from class j to class k at time $t+1$:

$$a_{jk} = \mathbb{P}\left\{\text{class } k \text{ at time } t+1 \mid \text{class } j \text{ at time } t\right\}. \tag{6}$$

Using Equation (6), $\boldsymbol{A}$ assumes a convenient probabilistic interpretation, which could be useful to answer many practical questions about relationships between classes. In Section 5, this is demonstrated on a case study where interactions between musical genres are analysed.

Importantly, a_{jk} in Equation (6) is also *not* constrained to be equal to a_{kj}. This introduces a notion of *directionality* between classes of items.

3.3 Parameter estimation

Jointly estimating γ and $\boldsymbol{A}$ is a complex task. Any change in the estimate of γ in any optimisation procedure would require an update to the EWMA distributions $\{\tilde{\boldsymbol{\pi}}_i^t\}$, which makes inference particularly cumbersome. One could, for example, use coordinate descent on an objective function of the pair $(\boldsymbol{A},\gamma)$, but inferring $\boldsymbol{A}$ for a fixed value of γ is also computationally expensive, which makes such an approach computationally impractical. Therefore, a two-step procedure is proposed. Instead of learning $\boldsymbol{A}$ and γ jointly, we suggest to first estimate γ, and subsequently estimate $\boldsymbol{A}$ conditional on the choice of the exploration parameter.

3.3.1 Selecting γ. A strategy for estimation of γ is proposed in this subsection. The EWMA distribution $\tilde{\boldsymbol{\pi}}_i^t$ should aim to approximate $\boldsymbol{\pi}_i^{t+1}$ closely, up to the effects of the transitions between classes, handled by $\boldsymbol{A}$ instead. To do this, we could minimise the negative log-likelihood (NLL) under the assumption that $\boldsymbol{A} = \boldsymbol{I}_N$, where $\boldsymbol{I}_N$ is the $N\times N$ identity matrix:

$$\gamma = \frac{1}{M}\sum_{i=1}^M \left\{ \underset{\gamma_i\in[0,1]}{\text{argmin}} \left[-\sum_{t=1}^{T-1}\sum_{j=1}^N n_{ij}^{t+1} \log\left(\tilde{\pi}_{ij}^t(\gamma_i)\right)\right]\right\}, \tag{7}$$

with initial value $\tilde{\pi}_{ij}^1 = (n_{ij}^1 + 1/N)/(\xi_i^1 + 1)$, which avoids the occurrence of the limit case $\log(0)$. In a Bayesian framework, such an initial value corresponds to the mean of the Dirichlet posterior distribution of the random variable corresponding to $\boldsymbol{\pi}_i^1$, assuming a Dirichlet conjugate prior with parameters set to $1/N$.

3.3.2 Estimating $\boldsymbol{A}$ conditional on γ. Conditional on the selected value of γ, the model parameters can be estimated by minimising the negative log-likelihood:

$$\hat{\boldsymbol{A}} = \underset{\boldsymbol{A}}{\text{argmin}} \left\{ -\sum_{t=1}^{T-1}\sum_{i=1}^M\sum_{j=1}^N n_{ij}^{t+1}\log\left(\sum_{k=1}^N \tilde{\pi}_{ik}^t a_{kj}\right)\right\}. \tag{8}$$

Note that the computational complexity for evaluating the objective function is $\mathcal{O}(N\sum_{t=1}^T\sum_{i=1}^M\sum_{j=1}^N \mathbb{I}\{n_{ij}^t > 0\})$, which is particularly convenient if the counts are sparse. To solve the optimisation procedure, we use the Adam algorithm [22].

3.4 Prediction

The performance of the PTM is assessed on its ability to predict future realisations of $\boldsymbol{n}_i^t$ and $\boldsymbol{\pi}_i^t$. It is therefore required to propose a prediction mechanism for those quantities. For an estimated transition matrix $\boldsymbol{A}$, the predicted frequency distribution of the classes at time $t+1$ is obtained as the conditional expectation

$\mathbb{E}(\boldsymbol{\pi}_i^{t+1} \mid \xi_i^{t+1}, \tilde{\boldsymbol{\pi}}_i^t)$:

$$\hat{\boldsymbol{\pi}}_i^{t+1} = \tilde{\boldsymbol{\pi}}_i^t \boldsymbol{A}. \tag{9}$$

To obtain a prediction for $\boldsymbol{n}_i^{t+1}$, we use the fact that $\mathbb{E}(\xi_i^{t+1} \mid \xi_i^t) = \xi_i^t$ from Equation (5), and the law of iterated expectation, which gives:

$$\hat{\boldsymbol{n}}_i^{t+1} = \xi_i^t \tilde{\boldsymbol{\pi}}_i^t \boldsymbol{A}. \tag{10}$$

A visual representation of the model, and the corresponding prediction procedure, is given in Figure 2.

4 EXPERIMENTS

In this section, we run experiments using different dynamic models for user preferences, demonstrating that the PTM achieves favorable performance on four different datasets.

4.1 Datasets

To show the wide applicability of the proposed modelling framework, we consider datasets from three different domains: music streaming, movie ratings, and restaurant recommendations.

Spotify. The streaming history of a sample of UK users of Spotify, a music streaming service. Each track is assigned to one of $N = 4430$ musical genres.[1]

Last.fm. The music streaming history of a sample of users on Last.fm[2] [8]. Each stream is assigned one of $N = 1500$ tags matched from the Million Song Dataset[3] [6].

MovieLens 25M. User-movie rating pairs[4] [18]. Each movie is assigned to a tag from a genome of $N = 1128$ tags.

Yelp. Restaurant reviews from the online review service Yelp.[5] Each restaurant is associated with one of $N = 192$ different cuisines.

The raw datasets are preprocessed to ensure that only users active in each time window are included in the analysis. Summary statistics for all preprocessed datasets are given in Table 1. For all datasets, we train models using the first $T - 1$ time periods, and we evaluate the predictive performance by using the last time period.

Table 1: Summary statistics of the datasets

Dataset	M	N	T	Interval	Start	End
Spotify	10000	4430	18	Quarter	Q1 2016	Q2 2020
Last.fm	450	1500	6	Quarter	Q3 2007	Q4 2008
MovieLens	1320	1128	5	Year	2015	2019
Yelp	1808	192	5	Year	2015	2019

4.2 Evaluation Tasks and Metrics

In this section, we introduce the four predictive tasks that we will use to evaluate the PTM and alternative methods. These tasks are depicted in Figure 3.

Total variation. The total variation between the observed class distribution $\boldsymbol{\pi}_i^t$ and its predicted value $\hat{\boldsymbol{\pi}}_i^t$ is:

$$v_i(t) = \frac{1}{2} \sum_{j=1}^{N} \left| \pi_{ij}^t - \hat{\pi}_{ij}^t \right|. \tag{11}$$

The minimum value of the score is 0, corresponding to identical distributions, whereas its maximum is 1. Therefore, the model should aim at minimising the total variation between the observed and predicted distributions. This task is evaluated using the average total variation (ATV) across all users, as represented in Figure 3a. The total variation task aims at evaluating the *global* performance of the model and its ability to predict the entire distribution of future allocations of user activity to classes.

Plus-minus (+/-). This task evaluates the ability of the model to predict whether the relative consumption of classes increases or decreases respectively from $\boldsymbol{\pi}_i^t$ to $\boldsymbol{\pi}_i^{t+1}$. The prediction problem can be framed as a simple binary classification problem, which is evaluated using the area under the receiver operating characteristic curve (ROC-AUC). This procedure is illustrated in Figure 3b. The task evaluates the performance of the model at estimating future consumption of items that the user has already interacted with.

New classes (t) *and* ***new classes (all).*** These tasks measure the model's ability to predict which new classes the user will interact with. Using a music streaming example, for the task *new class (t)*, a genre streamed at time $t+1$ is considered new if it was not streamed at t. On the other hand, in *new class (all)*, a genre is new only if it was *never* streamed by the user before time $t + 1$. This task can be again formulated as binary classification, and is evaluated using both ROC-AUC and PR-AUC (area under the precision-recall curve). PR-AUC is expected to more appropriately account for class imbalance. An illustration is given in Figure 3c. The two tasks evaluate the performance of the model at estimating which completely new classes of items the user might like.

4.3 Alternative Models and Baselines

We now describe alternative techniques for modelling user preferences over time that will serve as comparisons for the PTM. These are a mixture of state-of-the art methods and simple baselines.

Autoregressive Poisson model (Poisson AR). It is assumed that the counts $\boldsymbol{n}_i^t$ are distributed according to a Poisson distribution with rate depending on the observations at the previous time point, as follows: $\boldsymbol{n}_i^{t+1} \sim \text{Poisson}(\lambda)$ with $\lambda = \exp[\log(\boldsymbol{n}_i^t + 1)\boldsymbol{A}]$, where the exponential and logarithm are applied elementwise to the components of the vectors. The parameters are optimised via Adam using the negative log-likelihood loss function.

Dynamic Poisson factorisation (DPF) [9]. Users and classes are assigned K-dimensional time-dependent latent features $\boldsymbol{u}_i^t, \boldsymbol{v}_j^t \in \mathbb{R}^K$, evolving as Gaussian random-walks $\boldsymbol{u}_i^{t+1} \sim \mathcal{N}_K(\boldsymbol{u}_i^t, \sigma_u^2 \boldsymbol{I}_K)$ and $\boldsymbol{v}_j^{t+1} \sim \mathcal{N}_K(\boldsymbol{v}_j^t, \sigma_v^2 \boldsymbol{I}_K)$. Assuming baseline rates $\bar{\boldsymbol{u}}_i$ and $\bar{\boldsymbol{v}}_j$, counts are obtained as $n_{ij}^t \sim \text{Poisson}(\sum_{k=1}^K \exp\{u_{ik}^t + \bar{u}_{ik}\} \exp\{v_{jk}^t + \bar{v}_{jk}\})$. The model is fitted using code provided by the authors.[6] The two modeling features that separate the PTM from DPF are as follows. First, DPF assumes a low-rank representation of users and items,

[1] An interactive map of all the genres is publicly available at https://everynoise.com/.
[2] Available at http://ocelma.net/MusicRecommendationDataset/lastfm-1K.html.
[3] Available at: http://millionsongdataset.com/lastfm/.
[4] Available at: https://grouplens.org/datasets/movielens/25m/.
[5] Available at: https://www.yelp.com/dataset.
[6] Available at: https://github.com/blei-lab/DynamicPoissonFactorization.

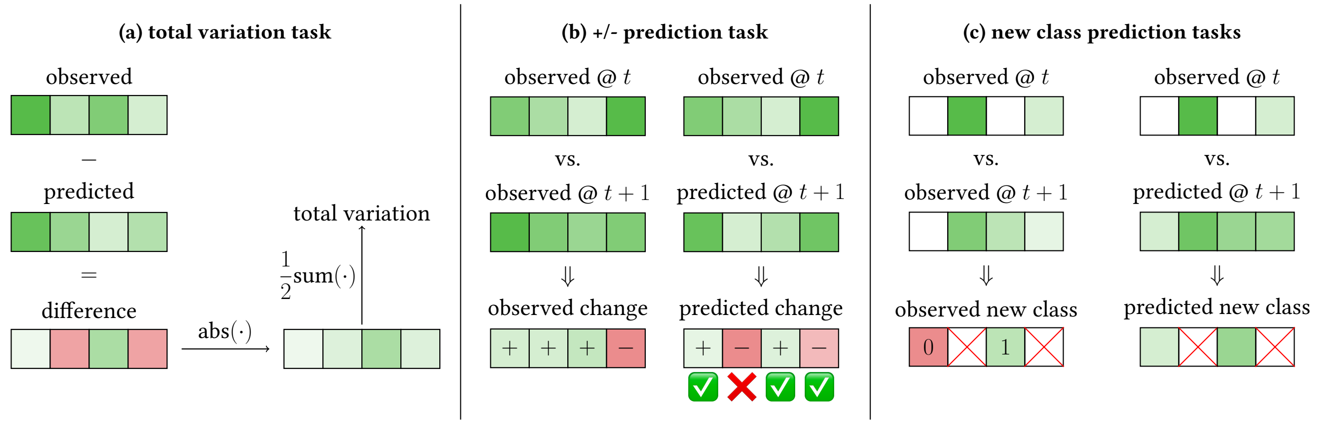

Figure 3: Visual representation of the prediction tasks. The color intensity corresponds to the consumption level of each class.

whereas the PTM does not. Second, DPF assumes that the evolution of users preferences are not predictable, and, therefore, they are modelled as a random walk. Consequently, in expectation, the latent features at time $t+1$ are identical to their values at the previous time window t, $\mathbb{E}(\boldsymbol{u}_i^{t+1} \mid \boldsymbol{u}_i^t) = \boldsymbol{u}_i^t$. In contrast, the PTM assumes that the dynamics governing the stochastic process of class counts are not completely random, and instead exhibit some structure.

Non-negative matrix factorization (NMF) (see, for example, [36, 40]). The $M \times N$ matrix of counts averaged over time is factorized into the dot product $\boldsymbol{W}\boldsymbol{H}^\top$ of two non-negative matrices $\boldsymbol{W} \in \mathbb{R}_+^{M\times K}$ and $\boldsymbol{H} \in \mathbb{R}_+^{N\times K}$ of rank K, obtained by minimisation of the Frobenius norm. The predicted count at the next time point is predicted as $\hat{n}_{ij}^{t+1} = \boldsymbol{w}_i^\top \boldsymbol{h}_j$, where the two vectors correspond to the i-th and j-th row of $\boldsymbol{W}$ and $\boldsymbol{H}$ respectively.

Previous observation. The predicted distribution of classes at time $t+1$ is simply the distribution at time t. Similarly, the predicted increase or decrease in the number of streams uses the difference observed between the two previous time periods. For prediction of new classes at time $t+1$ with respect to t, the number of interactions for each class at time $t-1$ is used. Observe that this is a special case of the PTM, with $\boldsymbol{A} = \boldsymbol{I}_N$ and $\gamma = 1$.

4.4 Predictive Performance

We apply every model to each of the four tasks, and we present the results in Table 2. For the PTM, the value of the exploration parameter γ is chosen using the NLL criterion in Equation (7), and is specified in parentheses in the table.

We observe that the PTM outperforms alternative models on all tasks and datasets, with the exception of prediction of new genres on *Last.fm*, and for total variation on *Yelp*. In particular, on *Last.fm*, DPF slightly outperforms the PTM at predicting new genres with respect to the entire user history, and the previous observation baseline achieves the best PR-AUC for prediction of new genres with respect to the previous time window. On *Yelp*, non-negative matrix factorization has a better performance than the PTM for the task of minimising total variation. We note that both *Last.fm* and *Yelp* are relatively small datasets, and thus they might favor simpler approaches with fewer parameters. On the Spotify dataset, which contains a few orders of magnitude more data, the PTM outperforms competing approaches much more significantly.

Inspecting the results, we find that modeling the changes in the *distribution* over items classes separately from the changes in overall *activity*—a distinctive feature of our model—is crucial in obtaining superior predictive performance. The PTM achieves this by the two-stage modelling approach given in Equation (5). The distribution of the classes $\boldsymbol{\pi}_i^t$ is modeled conditionally on the activity ξ_i^t, whereas all the alternative models attempt to infer the joint distribution, modelling $\boldsymbol{n}_i^t$ directly. The impact of this choice, independently of other modeling decisions, is best seen by comparing the Poisson AR and PTM models. Both models capture the evolution of preferences by using a transition matrix, but the former attempts to directly predict $\boldsymbol{n}_i^t$ whereas the latter predicts $\boldsymbol{\pi}_i^t$. Empirically, the latter works better for the datasets studied in this work. Even though the time periods are relatively long, we find that user activity is highly heterogeneous across different periods. In direct models for $\boldsymbol{n}_i^t$, the underlying effect of the relationships between classes appears to be confounded by the differences in activity levels across different time windows.

In summary, these results show that the PTM achieves favorable performance on various prediction tasks across multiple different datasets. Explicitly modeling preference trajectories thus appears to be beneficial when attempting to predict how tastes change over time. We believe that these improvements could translate to better recommender systems that anticipate users' changing needs.

However, explicitly modeling preference trajectories has a second major benefit: it allows us to directly inspect the structure of the preference graph and develop a qualitative understanding of the domain under consideration. In the next section, we use the Spotify dataset as a case study, and demonstrate how the PTM enables us to interpret and understand the dynamics of changing preferences in a real-world domain.

5 SPOTIFY CASE STUDY

In this section, we dive deeper into the Spotify dataset presented in Section 4.1. We begin by validating the two assumptions that drive

Table 2: Predictive performance of different models on four different datasets. The best result is highlighted in bold.

		Total variation	Plus-minus (+/-)	New classes (t)		New classes (all)	
Dataset	Model	TV	ROC-AUC	ROC-AUC	PR-AUC	ROC-AUC	PR-AUC
Spotify	PTM ($\gamma = 0.360$)	**0.378**	**0.698**	**0.944**	**0.267**	**0.889**	**0.039**
	Poisson AR	0.416	0.663	0.915	0.190	0.854	0.033
	DPF, $K = 5$	0.909	0.512	0.849	0.051	0.849	0.016
	NMF, $K = 50$	0.509	0.646	0.914	0.189	0.853	0.025
	Previous obs.	0.389	0.581	0.639	0.221	—	—
Last.fm	PTM ($\gamma = 0.803$)	**0.277**	**0.687**	**0.677**	0.421	0.655	0.183
	Poisson AR	0.320	0.646	0.648	0.420	0.600	0.174
	DPF, $K = 5$	0.616	0.591	0.659	0.393	**0.676**	**0.201**
	NMF, $K = 50$	0.294	0.684	0.669	0.416	0.607	0.150
	Previous obs.	0.301	0.591	0.634	**0.433**	—	—
MovieLens	PTM ($\gamma = 0.495$)	**0.601**	**0.685**	**0.800**	**0.269**	**0.790**	**0.134**
	Poisson AR	0.637	0.661	0.740	0.207	0.704	0.090
	DPF, $K = 5$	0.791	0.548	0.736	0.155	0.724	0.072
	NMF, $K = 50$	0.637	0.650	0.782	0.226	0.769	0.108
	Previous obs.	0.618	0.558	0.622	0.179	—	—
Yelp	PTM ($\gamma = 0.439$)	0.549	**0.716**	**0.864**	**0.326**	**0.834**	**0.145**
	Poisson AR	0.597	0.683	0.836	0.267	0.791	0.110
	DPF, $K = 5$	0.855	0.584	0.793	0.151	0.766	0.066
	NMF, $K = 50$	**0.542**	0.714	0.825	0.307	0.724	0.077
	Previous obs.	0.584	0.575	0.646	0.241	—	—

the development of our model. Then, we analyze the transition matrix $\boldsymbol{A}$ learned by the PTM, and we find that it contains meaningful and interpretable insights about the music space.

5.1 Dynamics and Diversity

As discussed in Section 1, the development of our model is driven by two observations: *a*) users' preferences change over time, and *b*) higher consumption diversity leads to stronger engagement. In this section, we validate these observations empirically.

Preferences Change Over Time. To show that the genres users listen to change over time, we proceed as follows. For each user, we compute the total variation distance $|\boldsymbol{\pi}_i^1 - \boldsymbol{\pi}_i^t|/2$ between their genre distribution during the first quarter in the dataset and that of every subsequent quarter $t = 2, \ldots, 18$. In Figure 4, we plot the histogram of distances (smoothed using a kernel density estimator) for each quarter. The figure shows that, on average, users progressively *shifted away* from the genres that they streamed in Q1 2016, and the difference appears to increase over time.

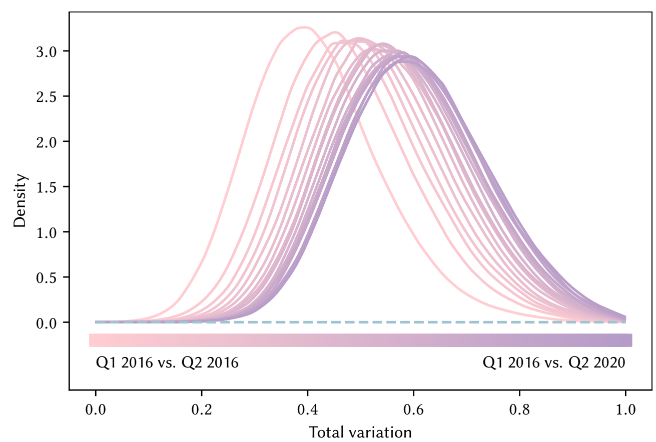

Figure 4: Estimated densities of total variation scores with respect to Q1 2016.

Diversity is Linked to Engagement. Next, we show that discovering new musical genres correlates with a user's propensity to convert to a paid subscription. To this end, we complement the dataset of Section 4 with another dataset containing 80,000 users that registered on Spotify before 2019. Every user in this dataset is continuously active on the platform between March 1st and September 30th, 2019, and is using a free (ad-supported) service during that period. Half of the users in our sample convert to a premium (paid subscription) service between October 1st and October 31st, 2019, whereas the other half of users do not convert. We partition each user's streams into two sets, those that occur between March 1st and August 31st, and those that occur between September 1st and September 30th. For each user, we look at the genres associated with every track that they stream, and we compute *a*) the fraction of distinct genres streamed for the first time during September, and *b*) the total number of distinct genres streamed during September. Figure 5 plots the average proportion of new genres against the total number of genres for the two aforementioned groups of users. For

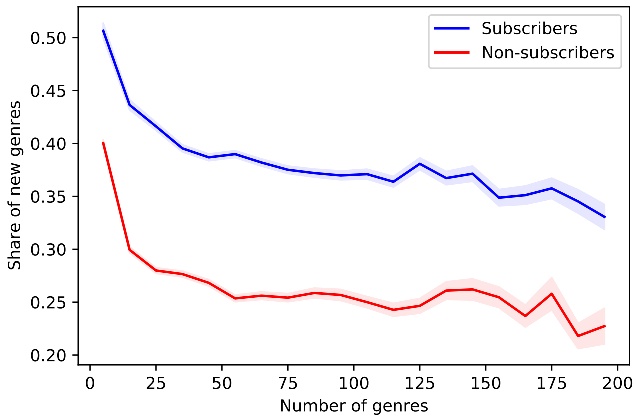

Figure 5: Proportion of new genres streamed in September 2019 (with respect to the period March 1st – August 31st) versus the total number of genres streamed in September 2019.

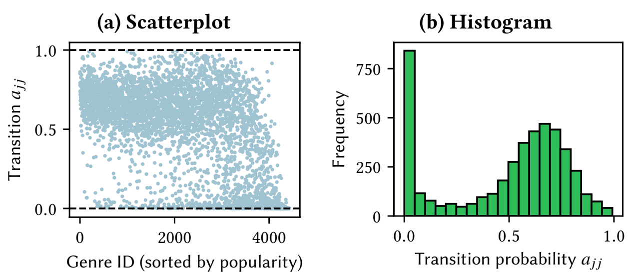

Figure 6: Scatterplot and histogram of the diagonal of $\boldsymbol{A}$.

every number of genres (i.e., activity level), users who converted to the paid subscription service (and thus, presumably, particularly enjoy the service) listened to a significantly larger fraction of new genres during September compared to users who did not convert. This suggests that helping users explore new regions of the music space can help increase users' engagement and satisfaction.

5.2 Structure in Preference Trajectories

Next, we inspect the transition matrix $\boldsymbol{A}$ learned by the PTM. One way to interpret this matrix is as the weighted adjacency matrix of a graph expressing a notion of distance between item classes. In the context of music streaming, we call this graph the *Genre Interaction Graph* (GIG).

First, we expect that the resulting graph has high-weight self-loops, represented by the diagonal entries in $\boldsymbol{A}$, since users are likely to continue listening, at time $t+1$, to the genres they were already listening to at time t. This is shown in Figure 6, which visualizes the distribution of the diagonal entries a_{jj} of $\boldsymbol{A}$, sorted by popularity. We find that users tend to stay in states corresponding to popular genres with high probability, but are more likely to leave states towards the low end of the popularity spectrum.

Now we turn to the off-diagonal elements of the GIG, which represent transitions between musical genres. By examining the structure of these transitions, it becomes possible to answer questions regarding the relationships between genres. For example:

(1) Given two genres i and j, what is the *best connecting genre*? In mathematical terms, we can answer this question by finding the genre $\hat{g}$ that maximises the sequential product of the transition probabilities:
$$\hat{g} = \underset{k\in\{1,\dots,N\}\setminus\{i,j\}}{\operatorname{argmax}} \quad a_{ik}a_{kj}. \tag{12}$$
(2) Given that the user streamed genre i, what is the most likely genre that the user will stream in the next time window? In this case, the solution is given by:
$$\hat{g} = \underset{k\in\{1,\dots,N\}\setminus\{i\}}{\operatorname{argmax}} \quad a_{ik}. \tag{13}$$
(3) What are the genres with the most asymmetric relationships? The directional nature of the PTM allows us to find genres i and j such that a_{ij} and a_{ji} are most different. We can rank pairs of genres by their absolute difference d_{ij}:
$$d_{ij} = |a_{ij} - a_{ji}|. \tag{14}$$

In Table 3, we report some examples obtained from the estimated GIG that answer these questions. For a pair of genres (i, j), we report the *best connecting genre* between i to j, and vice-versa. Also, we list the top genres that can be reached from i and j. The transitions in Table 3 are qualitatively meaningful and accord with intuitions. For example, the best way to connect the genre Spotify refers to as *Canadian pop* (e.g. Justin Bieber, Shawn Mendes and The Weeknd) to the *trap* sub-genre of rap (e.g. Migos, Ty Dolla $ign and Lil Uzi Vert) is estimated to be through *Toronto rap* (e.g. Drake, Tory Lanez and PARTYNEXTDOOR). As another example, *garage rock* (e.g. The Black Keys) might be linked to *Madchester* (e.g. Oasis or Joy Division) via *Sheffield indie* (e.g. the Arctic Monkeys). If a music streaming service or similar platform had a goal of helping users diversify their preferences, these connections would provide a useful guide.

The model captures many geographical characteristics, despite geography not being explicitly encoded in the PTM. For example, for two Italian genres in the table, the estimated transitions are only towards Italian music. Also, some of the most representative artists for the genres *indietronica* and *neo-psychedelic* are Australian (for the former, Rüfüs du Sol and Empire of the Sun, and for the latter, Tame Impala), which explains the link given by *Australian psych.*

Furthermore, for some of the estimated relationships the direction of the connection is important. For example, the best connecting genre between *pop* and *rap* is *pop rap* in both directions, but this is not the case for *indietronica* and *neo-psychedelic.* As pictorially explained in Figure 2, the PTM incorporates a notion of *directionality* between genres. To reinforce this point, in Table 4 we show the most asymmetric pairs of genres. In all cases, listeners of *genre 2* were far more likely to subsequently listen to *genre 1* than vice versa. In many cases, this is intuitive, and furthermore the most asymmetric relationships are often also hierarchical relationships. For example, *chanson* encapsulates *vintage chanson*, and *instrumental post-rock* and *polish post-rock* have a similar relationship. This suggests it would be possible to extract a hierarchy over genres using the GIG.

Table 3: Examples of transitions in the estimated genre interaction graph.

Genre 1	Genre 2	Path 1⟶2	Path 1⟵2	Top genres reachable from 1	Top genres reachable from 2
pop	rap	pop-rap	pop-rap	UK pop, dance pop, electropop	UK hip-hop, melodic rap, trap, ATL hip-hop
Canadian pop	trap	rap	Toronto rap	pop, rap, pop rap, R&B, Toronto rap	Toronto rap, ATL hip-hop, southern hip-hop
melodic rap	underground hip-hop	emo rap	trap	emo rap, vapor trap, DFW rap	UK alternative hip hop, vapor trap
garage rock	Madchester	Sheffield indie	Sheffield indie	modern alternative rock, Sheffield indie	Britpop, modern rock, rock, dance rock
moombahton	reggaeton	latin hip-hop	latin hip-hop	deep tropical house, electro house	latin, latin hip hop, trap latino, reggaeton flow
EDM	chillwave	progressive house	progressive house	tropical house, electro house, dance pop	indietronica, electronica, indie soul
indietronica	neo-psychedelic	alternative dance	Australian psych	electropop, vapor soul, indie pop	Australian psych, Edmonton indie, indie soul
Italian pop	Italian folk rock	classic Italian pop	Italian alternative	Italian arena pop, classic Italian pop	Italian alternative, Italian indie folk

Table 4: The most asymmetric relationships in the estimated genre interaction graph.

Genre 1	Genre 2	a_{12}	a_{21}
jam band	jamgrass	0.0014	0.87
Galician rock	Galician indie	0.0030	0.70
British choir	cathedral choir	0.0035	0.45
focus	jamgrass	0.0002	0.44
instrumental post-rock	Polish post-rock	0.0019	0.43
chanson	vintage chanson	0.0019	0.43

5.2.1 Most Likely Trajectories. Given two genres i and j, we can also use $\boldsymbol{A}$ to find a sequence of genres that connects i and j in such a way that every individual transition is likely. To this end, we adapt the criterion of Shahaf and Guestrin [34] and find a sequence $i, g_1, \ldots, g_K, j$ such that $\min_k a_{g_k g_{k+1}}$ is maximized. Additionally, we slightly penalize longer sequences by adding a regularization term to the objective.

We illustrate some of these trajectories in Figure 7. Given a set of *source* genres (in green) and a set of *destination* genres (in red), we plot the subgraph induced by the trajectories between every source and destination pair. Again, the resulting structure is highly intuitive. For example, starting from *classical* music, the *late romantic era* acts as a gateway to reach more contemporary genres of music, bridging centuries of musical history. Similarly, *tropical house*, an uplifting and relaxing type of electronic music, connects *edm* (electronic dance music) to a large part of the music space.

We envision that these trajectories could be turned into diversity-enhancing recommendations. Indeed, given a target genre that is far outside of a user's current preferences, we can use the graph to lay out a pathway that progressively brings them there, one manageable step at a time.

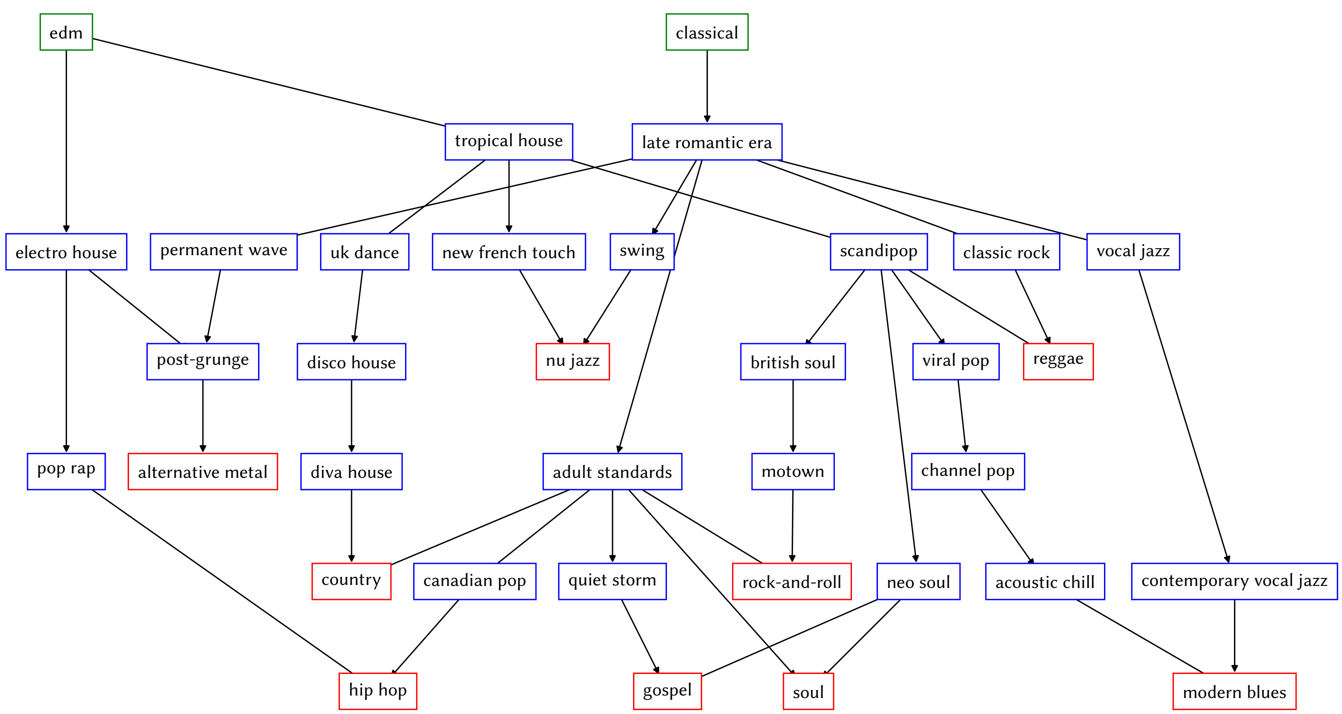

Figure 7: Graph induced by the most likely trajectories between every pair of source and destination nodes (in green and red, respectively). The trajectories are automatically learned from the PTM transition matrix A.

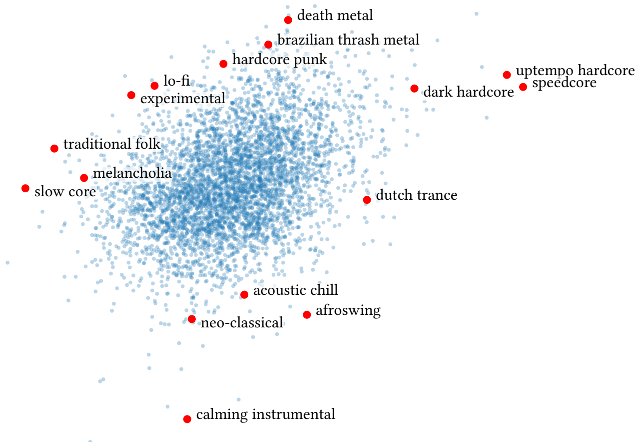

Figure 8: Projection of each row of the log **transition matrix onto two directions, intuitively corresponding to *tempo* (x-axis) and *energy* (y-axis).**

5.3 Visualizing the Item Space

Finally, the representation we learn can help us understand the macroscale structure of how genres relate to each other. We can interpret each row of $\log(\boldsymbol{A})$ (where the log is taken elementwise) as a N-dimensional representation of a genre in an Euclidean space. We project all genres onto two axes of meaning in music: *tempo* and *energy*. We operationalize these axes by defining dimensions as differences between the vector representations of a pair of genres. The *tempo* dimension is defined as *speedcore*−*slow core*, and the *energy* dimension is defined as *death metal*−*calming instrumental*. The resulting projection is shown in Figure 8. We observe that *tempo* and *energy* are significantly related, with a correlation coefficient of $r = 0.39$. Furthermore, we can position genres in a semantically meaningful space just from data on how users transition between them.

6 CONCLUSION

In this paper, we propose a dynamic model for user preferences towards items called Preference Transition Model (PTM). The model is well-suited to sequences of matrices of counts that express how many times users interacted with existing classes of items over time. A key feature of the model is that each user's activity in each time period, and how their activity is distributed across classes of items, is modelled separately. Empirically, we found this to be crucial to learn meaningful relationships between classes of items.

By comparing the PTM to alternative models and baselines on a number of different datasets, we showed that it outperforms all other methods. Furthermore, the model output—a matrix of transition probabilities between item classes—is highly interpretable, as demonstrated in a case study on the music streaming platform Spotify.

There are a number of promising directions for future work. The PTM in its current form does not scale well in the number of classes N, since the matrix $\boldsymbol{A}$ contains N^2 parameters. Low-rank approximations of $\boldsymbol{A}$ could be considered when the number of classes is large. The learned exploration parameters could also be parametrized by user features, e.g., age, to understand how the dynamics of preference change vary with user demographics and other traits. Also, in this work, the transition matrix $\boldsymbol{A}$ is assumed to be common across all users. In the music domain, previous research has shown that users could be efficiently classified in meaningful clusters based on their streaming habits [11]. In principle, this clustering structure could be used to assign group-specific transition matrices $\boldsymbol{A}$ or exploration parameters γ.

In summary, we have demonstrated that by explicitly modeling preference trajectories from raw user interaction data, we can achieve superior performance across a range of prediction tasks, as well as learn a useful, interpretable representation of how preferences change over time. This approach to user preference modeling could be valuable in settings where tastes change over time, as well as in the design of recommendation systems that suggest diverse and novel items in a principled way.

REFERENCES

[1] Zeinab Abbassi, Vahab S. Mirrokni, and Mayur Thakur. 2013. Diversity Maximization under Matroid Constraints. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '13)*. Association for Computing Machinery, New York, NY, USA, 32–40.

[2] Ayan Acharya, Joydeep Ghosh, and Mingyuan Zhou. 2015. Nonparametric Bayesian Factor Analysis for Dynamic Count Matrices *(Proceedings of Machine Learning Research, Vol. 38)*, Guy Lebanon and S. V. N. Vishwanathan (Eds.). PMLR, San Diego, California, USA, 1–9.

[3] Gediminas Adomavicius and Alexander Tuzhilin. 2005. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering* 17, 6 (2005), 734–749.

[4] Ashton Anderson, Lucas Maystre, Ian Anderson, Rishabh Mehrotra, and Mounia Lalmas. 2020. Algorithmic Effects on the Diversity of Consumption on Spotify. In *Proceedings of The Web Conference 2020* (Taipei, Taiwan) *(WWW '20)*. Association for Computing Machinery, New York, NY, USA, 11 pages.

[5] Andrew Arnold, Yan Liu, and Naoki Abe. 2007. Temporal Causal Modeling with Graphical Granger Methods. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Jose, California, USA) *(KDD '07)*. Association for Computing Machinery, New York, NY, USA, 66–75.

[6] Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman, and Paul Lamere. 2011. The Million Song Dataset. In *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR 2011)*.

[7] Alex Beutel, Paul Covington, Sagar Jain, Can Xu, Jia Li, Vince Gatto, and Ed H. Chi. 2018. Latent Cross: Making Use of Context in Recurrent Recommender Systems. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* (Marina Del Rey, CA, USA) *(WSDM '18)*. Association for Computing Machinery, New York, NY, USA, 46–54.

[8] O. Celma. 2010. *Music Recommendation and Discovery in the Long Tail*. Springer.

[9] Laurent Charlin, Rajesh Ranganath, James McInerney, and David M. Blei. 2015. Dynamic Poisson Factorization. In *Proceedings of the 9th ACM Conference on Recommender Systems* (Vienna, Austria) *(RecSys '15)*. Association for Computing Machinery, New York, NY, USA, 155–162.

[10] Munmun De Choudhury, Scott Counts, and Mary Czerwinski. 2011. Identifying Relevant Social Media Content: Leveraging Information Diversity and User Cognition. In *Proceedings of the 22nd ACM Conference on Hypertext and Hypermedia* (Eindhoven, The Netherlands) *(HT '11)*. Association for Computing Machinery, New York, NY, USA, 161–170.

[11] Paolo Dragone, Rishabh Mehrotra, and Mounia Lalmas. 2019. Deriving User- and Content-Specific Rewards for Contextual Bandits. In *The World Wide Web Conference* (San Francisco, CA, USA) *(WWW '19)*. Association for Computing Machinery, New York, NY, USA, 2680–2686.

[12] Daniel M. Dunlavy, Tamara G. Kolda, and Evrim Acar. 2011. Temporal Link Prediction Using Matrix and Tensor Factorizations. *ACM Trans. Knowl. Discov. Data* 5, 2, Article 10 (Feb. 2011), 27 pages.

[13] Michael Eichler, Rainer Dahlhaus, and Johannes Dueck. 2017. Graphical Modeling for Multivariate Hawkes Processes with Nonparametric Link Functions. *Journal of Time Series Analysis* 38, 2 (2017), 225–242.

[14] Jalal Etesami, Negar Kiyavash, Kun Zhang, and Kushagra Singhal. 2016. Learning Network of Multivariate Hawkes Processes: A Time Series Approach. In *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence* (Jersey City, New Jersey, USA) *(UAI'16)*. AUAI Press, Arlington, Virginia, USA,

162–171.
[15] Daniel Fleder and Kartik Hosanagar. 2009. Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity. *Management Science* 55, 5 (2009), 697–712.
[16] Prem Gopalan, Jake M. Hofman, and David M. Blei. 2015. Scalable Recommendation with Hierarchical Poisson Factorization. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence* (Amsterdam, Netherlands) *(UAI'15)*. AUAI Press, Arlington, Virginia, USA, 326–335.
[17] S. Gultekin and J. Paisley. 2014. A Collaborative Kalman Filter for Time-Evolving Dyadic Processes. In *2014 IEEE International Conference on Data Mining*. 140–149.
[18] F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. *ACM Trans. Interact. Intell. Syst.* 5, 4, Article 19 (2015).
[19] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. [n.d.]. Session-based Recommendations with Recurrent Neural Networks. In *4th International Conference on Learning Representations, ICLR 2016, year = 2016*, Yoshua Bengio and Yann LeCun (Eds.). San Juan, Puerto Rico.
[20] David Holtz, Ben Carterette, Praveen Chandar, Zahra Nazari, Henriette Cramer, and Sinan Aral. 2020. The Engagement-Diversity Connection: Evidence from a Field Experiment on Spotify. In *Proceedings of the 21st ACM Conference on Economics and Computation* (Virtual Event, Hungary) *(EC '20)*. Association for Computing Machinery, New York, NY, USA, 75–76.
[21] Alexandros Karatzoglou, Xavier Amatriain, Linas Baltrunas, and Nuria Oliver. 2010. Multiverse Recommendation: N-Dimensional Tensor Factorization for Context-Aware Collaborative Filtering. In *Proceedings of the Fourth ACM Conference on Recommender Systems* (Barcelona, Spain) *(RecSys '10)*. Association for Computing Machinery, New York, NY, USA, 79–86.
[22] Diederik P. Kingma and Jimmy Ba. 2015. Adam: a Method for Stochastic Optimization. In *3rd International Conference on Learning Representations, ICLR*, Yoshua Bengio and Yann LeCun (Eds.). San Diego, CA, USA.
[23] Yehuda Koren. 2009. Collaborative Filtering with Temporal Dynamics. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Paris, France) *(KDD '09)*. Association for Computing Machinery, New York, NY, USA, 447–456.
[24] Matevž Kunaver and Tomaž Požrl. 2017. Diversity in recommender systems – A survey. *Knowledge-Based Systems* 123 (2017), 154–162.
[25] Bin Li, Xingquan Zhu, Ruijiang Li, Chengqi Zhang, Xiangyang Xue, and Xindong Wu. 2011. Cross-Domain Collaborative Filtering over Time *(IJCAI'11)*. AAAI Press, 2293–2298.
[26] Scott W. Linderman and Ryan P. Adams. 2014. Discovering Latent Network Structure in Point Process Data. In *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32* (Beijing, China) *(ICML'14)*. II–1413–II–1421.
[27] Hao Ma, Xueqing Liu, and Zhihong Shen. 2016. User Fatigue in Online News Recommendation. In *Proceedings of the 25th International Conference on World Wide Web*. Montreal, QC, Canada.
[28] B. Mark, G. Raskutti, and R. Willett. 2019. Network Estimation From Point Process Data. *IEEE Transactions on Information Theory* 65, 5 (2019), 2953–2975.
[29] Julian John McAuley and Jure Leskovec. 2013. From Amateurs to Connoisseurs: Modeling the Evolution of User Expertise through Online Reviews. In *Proceedings of the 22nd International Conference on World Wide Web* (Rio de Janeiro, Brazil) *(WWW '13)*. Association for Computing Machinery, New York, NY, USA, 897–908.
[30] Tien T. Nguyen, Pik-Mai Hui, F. Maxwell Harper, Loren Terveen, and Joseph A. Konstan. 2014. Exploring the Filter Bubble: The Effect of Using Recommender Systems on Content Diversity. In *Proceedings of the 23rd International Conference on World Wide Web* (Seoul, Korea) *(WWW '14)*. Association for Computing Machinery, New York, NY, USA, 677–686.
[31] Eli Pariser. 2011. *The Filter Bubble: What the Internet Is Hiding from You.*
[32] S.W. Roberts. 1959. Control Chart Tests Based on Geometric Moving Averages. *Technometrics* 1, 3 (1959), 239–250.
[33] Aaron Schein, John Paisley, David M. Blei, and Hanna Wallach. 2015. Bayesian Poisson Tensor Factorization for Inferring Multilateral Relations from Sparse Dyadic Event Counts. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Sydney, NSW, Australia) *(KDD '15)*. Association for Computing Machinery, New York, NY, USA, 1045–1054.
[34] Dafna Shahaf and Carlos Guestrin. 2010. Connecting the Dots between News Articles. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Washington, DC, USA) *(KDD '10)*. Association for Computing Machinery, New York, NY, USA, 623–632.
[35] Johan Ugander, Lars Backstrom, Cameron Marlow, and Jon Kleinberg. 2012. Structural diversity in social contagion. *Proceedings of the National Academy of Sciences* 109, 16 (2012), 5962–5966.
[36] Y. Wang and Y. Zhang. 2013. Nonnegative Matrix Factorization: A Comprehensive Review. *IEEE Transactions on Knowledge and Data Engineering* 25, 6 (2013), 1336–1353.
[37] Chao-Yuan Wu, Amr Ahmed, Alex Beutel, Alexander J. Smola, and How Jing. 2017. Recurrent Recommender Networks. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining* (Cambridge, United Kingdom) *(WSDM '17)*. Association for Computing Machinery, New York, NY, USA, 495–503.
[38] Liang Xiong, Xi Chen, Tzu-Kuo Huang, Jeff Schneider, and Jaime G. Carbonell. 2010. *Temporal Collaborative Filtering with Bayesian Probabilistic Tensor Factorization.* 211–222.
[39] Mi Zhang and Neil Hurley. 2008. Avoiding Monotony: Improving the Diversity of Recommendation Lists. In *Proceedings of the 2008 ACM Conference on Recommender Systems (RecSys '08)*. Association for Computing Machinery, New York, NY, USA, 123–130.
[40] Sheng Zhang, Weihong Wang, James Ford, and Fillia Makedon. 2006. *Learning from Incomplete Ratings Using Non-negative Matrix Factorization.* 549–553.