MINING ONLINE MUSIC LISTENING TRAJECTORIES

Flavio Figueiredo[1] **Bruno Ribeiro**[2] **Christos Faloutsos**[3]
Nazareno Andrade[4] **Jussara M. Almeida**[5]
[1] IBM Research - Brazil [2] Purdue University [3] Carnegie Mellon University
[4] Universidade Federal de Campina Grande [5] Universidade Federal de Minas Gerais

ABSTRACT

Understanding the listening habits of users is a valuable undertaking for musicology researchers, artists, consumers and online businesses alike. With the rise of Online Music Streaming Services (OMSSs), large amounts of user behavioral data can be exploited for this task. In this paper, we present SWIFT-FLOWS, an approach that models user listening habits in regards to how user attention transitions between artists. SWIFT-FLOWS combines recent advances in trajectory mining, coupled with modulated Markov models as a means to capture both how users switch attention from one artist to another, as well as how users fixate their attention in a single artist over short or large periods of time. We employ SWIFT-FLOWS on OMSSs datasets showing that it provides: (1) semantically meaningful representation of habits; (2) accurately models the attention span of users.

1. INTRODUCTION

Is it possible to create expressive yet succinct representations of individuals' music listening habits? Are there common patterns on how music is listened to across different genres and different artists that have highly different popularity? For a long time such questions have attracted the attention of researchers from different fields. In the fields of psychology and musicology [10, 20, 21], researchers exploit musical preferences to study social and individual identity [20], mood regulation [23], as well as the underlying factors of preferences [21]. Computer scientists are also tackling such questions as they become central to develop music recommender systems [3, 4, 7].

With the rise of Online Music Streaming Services (OMSSs) over the last decade, large datasets of user [1] behavior can be used to shed light on questions like the ones above. More specifically, digital traces of the listening habits of individuals are readily available to researchers.

[1] Since our case study is on Online Music Streaming Services (OMSSs), we use the terms users and listeners interchangeably.

In this paper, we focus on the online listening habits of users as trajectories [7] (or trails [24]). Given that a user, u, listens to music by switching attention between different artists, a trajectory captures the sequence of artists or songs visited by a user when listening to music. The main contribution of this paper is to present the SWIFT-FLOWS [2] model, a general technique designed to study user trajectories in OMSSs. We tackle several challenges that stem from the complexity of user behavior, such as:

(a) *Asynchronous users with mixed but similar behavior*: Users that consume music from a set of artists will not start their playlists at the same time or listen to songs in the same order.
(b) *Repeated consumption*: Users tend to listen to artists in bursts, more than what one would expect at random in a shuffled playlist.
(c) *Biased Observations & Small Subpopulations*: User behavior datasets are naturally sparse and biased towards more popular artists. Nevertheless, we still want to be able to analyze underrepresented subpopulations of users and artists.

SWIFT-FLOWS effectiveness is evaluated in large datasets, with results showing that SWIFT-FLOWS: (1) captures semantically meaningful representation of artist transitions; (2) accurately models the attention span of users.

2. RELATED WORK

Understanding the listening habits of individuals has attracted interest from different research fields. Among other problems, musicologists and social psychologists have looked into the latent factors that explain musical preferences [20, 21], factors that affect listener experience (e.g., Music itself, Situational Factors and the Listener him/herself) [10], as well as the relationships between musical imagination and human creativity [10].

Regarding the material methods listeners exploit to listen to music, Nowak [16] discussed the social-material relations of music consumption. The authors conclude that even the same user still relies on multiple forms of listening to music (e.g., legal and illegal downloading, streaming services, CDs, etc). These various forms of consumption were also discussed by Bellogin *et al.* [1]. Here, the au-

[2] Switch and Fixation Trajectory Flows

thors showed the disagreement between different web and social music services (in terms of artist popularity).

Several studies, such as the ones Marques *et al.* [12], Park *et al.* [18], and Moore *et al.* [15], characterized the exploratory behavior of users in OMSSs. Marques *et al.* [12] described the habits of users when searching for novel songs to listen to. Park *et al.* [18] defined a new measure to compute the diverseness of musical tastes. Moore *et al.* [15] looked into the tastes of users over time through the use of Latent Markov Embedding (LME) [3,4].

In contrast with the aforementioned studies, our work with SWIFT-FLOWS is focused on extracting the latent trajectories that explain user attention when listening to music online. Nevertheless, SWIFT-FLOWS can be used to tackle problems as the ones described. For instance, we are able to aggregate the preferences of users from different demographics as shown in Section 4. For musicologists and psychologists, these results indicate how SWIFT-FLOWS can be used as a tool to better understand the hidden factors that define our consumption habits. Regarding OMSSs, previous work [12, 17, 18] usually relied on defining specific point estimates that are used to understand and capture listening behavior. Such measures are susceptible to effects (a-c) described in Section 1.

To capture both the inter-artist transitions, those were a listener changes attention from one artist to another, as well the long and short tails of listener fixation in a single artist, SWIFT-FLOWS advances the state-of-the-art [7] by defining a combined approach that can capture both behaviors. Inter-artist transitions, or switches in attention, are captured by exploring the ideas in [7]. Intra-artist transitions, or fixation, is captured by modulated Markov models [22]. In this sense, SWIFT-FLOWS provides a interpretable results than [7] or [22] in isolation. We describe the details of the model next.

3. THE SWIFT-FLOWS MODEL

We now describe SWIFT-FLOWS. Let $\mathcal{D}$ be a dataset consisting of $(\text{user}, \text{artist}, \text{timestamp})$ tuples observed over a time window (i.e., the temporal range of our datasets). Each tuple registers that a user listened to songs from an artist at a moment in time. Let $u \in \mathcal{U}$ define the set of users and $a \in \mathcal{A}$ define the set of artists. By ordering $\mathcal{D}$ according to the timestamps, each user can be represented as a trajectory: $T_u =< a_{u,1}, a_{u,2}, ..., a_{u,|\mathcal{T}_u|} >$. This trajectory represents the history of the user listening to music transitioning between songs of a same artist – in intra-artist ($a_{u,i} = a_{u,i+1}$) transitions – and songs from different artists – in inter-artist ($a_{u,i} \neq a_{u,i+1}$) transitions.

Both inter and intra artist transitions are important when studying trajectories. Inter-artist transitions capture a switch in users attention from one artist to another, whereas intra-artist transitions captures a fixation on a same artist. SWIFT-FLOWS isolates both effects and exploits stochastic complementation [14] to propose two complementary Markov models, as illustrated in Figure 1, that together are able to capture both the intra-artist and inter-artist transition behavior. Isolation of intra from inter artist transitions is necessary to model both the long and and short attention tails of repeated consumption [6, 22].

The *intra-artist* (Figure 1-b), or fixation, model consists of a modulated Markov model that is able to capture how users revisit artists. Intra/inter transition separation is possible by treating user attention as a reducible system, where we model the strong memory of intra-artist transitions – some users continuously listen to the same artist for hours – as only interfering with the inter-artist dynamics through limited user attention. This creates an effective separation between the intra-artist model and the inter-artist model. A play takes us to an inter-artist transition from artist s to artist d, $s \neq d$, which then again transitions to the intra-artist model of artist d. The *inter-artist* (Figure 1-c) attention, or switch, transitions are captured by a graphical model, using a Bayesian approach to estimate inter-artist transitions. This approach avoids problems associated with point estimates [7, 19] and is robust to infrequent transitions of small sub-populations of interest.

Data Representation: Let users (artists) to be numbered between one and $|\mathcal{U}|$ ($|\mathcal{A}|$). Let n_{dsu}, the number of times user $u \in \mathcal{U}$ transitioned from $s \in \mathcal{A}$ to $d \in \mathcal{A}$:

$$n_{dsu} = \sum_{i=2}^{|T_u|} \mathbf{1}(a_{u,i-1} = s \wedge a_{u,i} = d), \tag{1}$$

where, $\mathbf{1}$ is an indicator function that will evaluate to 1 when $a_{u,i-1} = s$ and ($\wedge$) $a_{u,i} = d$, 0 otherwise.

With these counts, we can define a tensor $\boldsymbol{\mathcal{X}}$ (as shown in Figure 1-a) $\boldsymbol{\mathcal{X}} = \left[\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_{|\mathcal{U}|}\right]$, where $\mathbf{X}_u$ is:

$$\mathbf{X}_u = \begin{bmatrix} n_{11u} & \cdots & n_{1|\mathcal{A}|u} \\ \vdots & \ddots & \vdots \\ n_{|\mathcal{A}|1u} & \cdots & n_{|\mathcal{A}||\mathcal{A}|u} \end{bmatrix} \tag{2}$$

This data representation is distinct from other tensor decompositions that mine $\mathcal{D}$ in its original "user", "object" and "time" coordinates as the three tensor modes [13, 25]. These techniques are meant to capture synchronous user behavior. As shown in previous work [7], the representation of $\boldsymbol{\mathcal{X}}$ is more suitable to capture the asynchronous but similar behavior patterns that emerge when we have a mixed population of users, spread across different time zones and with different activity patterns as in OMSSs.

We now describe both the inter-artist an intra-artist models. In the following, we use the "$\cdot$" notation to imply a sum over a given dimension (e.g., $n_{ds\cdot} = \sum_{u \in \mathcal{U}} n_{dsu}$).

3.1 Switch Model

To model inter-artist transitions, we define $\boldsymbol{\mathcal{X}}^- = \boldsymbol{\mathcal{X}} - diagonals(\boldsymbol{\mathcal{X}})$ by removing the cases where $s = d$ from $\boldsymbol{\mathcal{X}}$, since this behavior is captured by the Fixation model (next subsection). Our goal with the Switch model is to estimate trajectories as an interpretable probability space. That is, our goal is to decompose $\boldsymbol{\mathcal{X}}$ in a probability matrix $\mathbf{P}$, where each cell in this matrix captures the probability of a user switching attention for s to d (or $p(d|s)$).

A naïve way to define $\mathbf{P}$ is simply to define $p(d|s) \propto n_{ds\cdot}$. That is, to use maximum likelihood estimates [11].

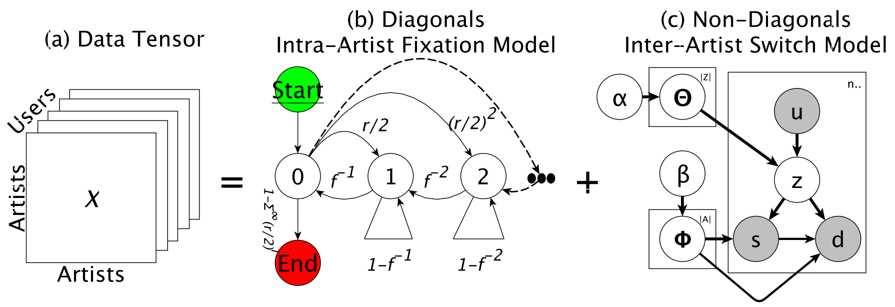

Figure 1. The SWIFT-FLOWS model: Data representation by tensor $\boldsymbol{\mathcal{X}}$ (left), the repeated consumption model (center) and the inter-artist graphical model (right).

However, this approach has three undesirable properties [19]: (1) there are not enough samples to accurately estimate the transition probabilities for most artists; (2) the transition probability matrix $\mathbf{P}$ is sparse, stating that it is impossible to transition from an artist s to d when no user has done so in the past; and, (3) it does not take into account user preferences. For example, if we observe that the transition Sepultura→Beyonce is very common for a single or small group of users, this does not imply that it is frequent for all of the listeners in the OMSSs.

In order to deal with such issues, we employ the Bayesian model depicted in Figure 1-c. With this model, our goal is capture the latent spaces of inter-artist transition patterns shared by a group of users. We call this the Switch model. The latent space $\mathcal{Z}$ defines a set of *transitions* between pairs of artists s and d. We refer to each latent factor z as an *attention transition gene*, and the collection of genes as a *genome*. These terms are inspired by the "Music Genome Project", a proprietary approach developed by Pandora that aims to describe in detail the musical characteristics of individual songs [3] .

Estimating the Model: Let $k = |\mathcal{Z}|$ ($k << |\mathcal{A}|$) be an input variable determining the number of genes (or latent factors) to be estimated. Later, we shall describe our approach to define k. The two other inputs are the hyper-parameters α and β. The outputs of the Switch model are two matrices, $\mathbf{\Theta}$ and $\mathbf{\Phi}$, as well as a vector $\mathbf{z}$. $\mathbf{\Theta}$ has $|\mathcal{U}|$ rows and $|\mathcal{Z}|$ columns, where each cell contains the probability that a user has a preference towards a given gene:

$$p(z|u) = \mathbf{\Theta}(u, z) = \theta_{z|u}(z) = \frac{n_{zu} + \alpha}{n_{\cdot u} + |\mathcal{Z}|\alpha} \quad (3)$$

where n_{zu} is estimated by the model. Matrix $\mathbf{\Phi}$ has $|\mathcal{Z}|$ rows and $|\mathcal{A}|$ columns. It captures the probability that when a user is interest in gene z it will transition to a, i.e.:

$$p(a|z) = \mathbf{\Phi}(z, a) = \phi_{a|z}(a) = \frac{n_{az} + \beta}{n_{\cdot z} + |\mathcal{A}|\beta} \quad (4)$$

where, once again, n_{az} is estimated from the data by the model. Finally, vector $\mathbf{z}$ contains the probabilities of each gene $z \in \mathcal{Z}$, referred to as $p(z)$, that is: $p(z) \propto n_z$. Finally, the decomposed transition matrix $\mathbf{P}$ is defined by:

$$\mathbf{P}(s, d) = \sum_{z \in |\mathcal{Z}|} p(z|s)p(d|z) \quad (5)$$

where $p(d, s|z) = p(s|z)p(d|z)$, and $p(z|s) \propto p(s|z)p(z)$.

Gibbs Sampling: We use a collapsed Gibbs sampler [8] to estimate matrices $\mathbf{\Theta}$ and $\mathbf{\Phi}$ by estimating n_{zu} and n_{az}, as well as vector $\mathbf{z}$. We sample from the posterior defined by the product $\theta_{z|u}\phi_{s|z}\phi_{d|z}$ [7]. We fix hyper-parameters $\alpha = \frac{50}{|\mathcal{Z}|}$, and $\beta_s = \beta_d = 0.001$, as is usually done with similar models [7, 13]. We execute the sampler for 800 iterations with 300 being discarded as burn-in.

Estimating k: We apply the minimum description length (MDL) principle [9], which is largely used for problems of model selection, to determine the number of genes $k = |\mathcal{Z}|$. With MDL, we fine tune SWIFT-FLOWS in order to extract a succinct, yet still accurate, representation of the listening habits of users. MDL captures how good a model $\mathcal{M}$ ($\mathbf{P}$ in our case) represents the data by taking into account the trade-off between the "goodness" (or likelihood) and the complexity (or generality) of the model.

To apply MDL we first define the likelihood of the data given the model $\mathcal{M}$. Given $n_{ds} = n_{\cdot ds}$ the number of transitions from s to d by all users, the log likelihood of matrix $\mathbf{P}$ is given by $\sum_{s,d|s \neq d} n_{ds} log(p(d|s))$ [4] . The MDL cost of model $\mathcal{M}$ is given by the sum:

$$Cost(\mathbf{P}, \mathcal{M}) = Cost(\mathbf{P} \mid \mathcal{M}) + Cost(\mathcal{M}). \quad (6)$$

$Cost(\mathbf{P} \mid \mathcal{M})$, defined as the negative log-likelihood, captures the goodness-of-fit of the data given the model: higher-values imply on accurate but yet succinct (less factors) recoveries of $\mathbf{P}$. $Cost(\mathcal{M})$ captures the complexity:

$$Cost(\mathcal{M}) = log^*(|\mathcal{A}|) + log^*(|\mathcal{Z}|) + \sum_{s,d,z} [log^*(\lceil p(d|z) n_{\cdot\cdot} \rceil) + log^*(\lceil p(s|z) n_{\cdot\cdot} \rceil) + log^*(\lceil p(z) n_{\cdot\cdot} \rceil)]$$

where $log*$ is the universal coding cost (number of bits) for integers [9]. $Cost(\mathcal{M})$ represents the encoding each matrix in the model in integer representation with precision $n_{\cdot\cdot}$ (the total number of transitions) [5] .

[3] `http://www.pandora.com/about/mgp`

[4] The likelihood is the product of $p(d|s)$ for all n_{ds} transitions [5].

[5] Since we deal with counts, the smallest probability value is $(1/n_{\cdot\cdot})$.

3.2 Fixation Model

Users' bursty repeated consumption of artists requires modeling this behavior with a stochastic process that has memory. Markov modulated processes are a class of models that are particularly versatile for this task [22]. Our goal here is not only to model user behavior but also, through the use of intuitive parameters, understand how users repeatedly consume artists. Most importantly, we want to reproduce user attention giving rise to both exponential and power law distributions observed in our datasets.

Our fixation model, which captures the intra-artist transitions, is a Markov modulated process where we use an infinite number of states, an approach widely used to model systems with bursty behavior [22]. Figure 1 (b) illustrates our model (only the initial states). The "start" circle represents the initial transition from the Inter-artist model. From state zero we are interested in how long it takes to exit from the "exit" transition. Thus, circles "start" and "exit" in Figure 1 (b) are not states but rather entrance (exit) transitions from (to) the inter-artist model. The states of the model capture the affinity of the user for the artist, that is how much the user is willing to repeatedly listen the artist's songs. There is a fixed residency time Δt on each state. Thus, higher states represent that the user has a higher affinity and thus dedicates more play-time to the artist.

The model has parameters $0<r<1$ and $1\leq f<4r$. The limit of $4r$ is required as described in [22] [6]. Parameter r models the user "rush", capturing how users are led to hear more from an artist (e.g. entire album) after hearing some music by this artist. Parameter f models user "fixation", representing how long it takes for users to get over an initial impulse to listen to an artist, which is also a function of the artist's song inventory size. A large value of f implies that users quickly get over their initial impulse or happens because the artist has just a few songs.

We can fit the Fixation model to the complementary cumulative distribution function (CCDF) of the time users dedicate to an artist using the Levenberg-Marquardt algorithm. The CCDF will define the probability of the residency time in the chain. The infinite number of states can be captured by using a sufficient number of states (100 in our datasets). We evaluate the algorithm on the mean squared error of the real data and the residency times generated by the model. As we shall show empirically in our results, this model is capable of generating both power-law and exponential residency times as also discussed in [22].

One interesting property of the Fixation model is that it is able to estimate the expected amount time users will fixate on a given artist. To achieve this, we can compute the expected number of steps that it takes to go from the Start state to the End state [11] [7]. If we define this value per artist as e_a, we can couple the Switch model with the Fixation model by estimating the expected fixation steps per latent spaces, e_z, as:

$$e_z = \sum_{a \in |\mathcal{A}|} p(a|z) e_a \tag{7}$$

That is, e_a is the expected number of steps a user will remain in artist a with regards to his/her interest in gene z.

In the next section we describe SWIFT-FLOWS at work.

4. SWIFT-FLOWS AT WORK

We apply SWIFT-FLOWS on datasets crawled from Last.FM. Last.FM aggregates various forms of digital music consumption, ranging from desktop/mobile media players to streaming services (theirs and others) [8]. Last.FM is also an online social network (OSN), allowing the creation of user groups as well as providing demographical data. The datasets we explore are:

Last.FM-2009 Collected in using a snowball sampling [2]. After the snowball sampling, 992 uniformly random users were selected. The dataset contains, for each user, the complete listening history (all plays) from February 2005 to May 2009, the self-declared nationality, age (at the time), and registration date [2]. This dataset accounts for 18.5 million user, artist, and timestamp triples, as well as 107,397 unique artists.

Last.FM-2014 Crawled in 2014 by identifying users that participate in discussion groups on Last.FM. Contains the listening history (from February 2005 to August 2014) of a subset of the users that discuss pop-artists on Last.FM discussion groups. The total number of users in this dataset is 15,329. Also, this dataset contains 836,625 unique artists and roughly 218 million user, artist, and timestamp triples. As is the case with Last.FM-2009, this dataset provides the age and nationality of all the users.

Because of these various means of consumption [16], Last.FM presents itself as an interesting platform for studying *online behavior*. The service aggregates user accesses from desktop media players (that incorporates legal and illegal downloads), free, and also paid streaming services. Nevertheless, it is important to point out that the observed attention trajectories will be impacted by how the data was gathered (e.g., Last.FM-2014 has a bias towards pop artists), as well as the internal mechanisms of the OMSSs (e.g., such as recommendation services and user interfaces). As we shall discuss in our results, regardless of the data biases, SWIFT-FLOWS is able to represent the attention trajectories of under-represented user populations.

We run the inter-artist attention Switch model of SWIFT-FLOWS on $\boldsymbol{\mathcal{X}}^{-}$, and the Fixation model of SWIFT-FLOWS on the intra-artist transitions. In both cases, only artists which had at least five plays by five users are considered. In total, the Last.FM-2014 dataset has over 3M plays of such artists, while Last.FM-2009 has roughly

[6] The authors write the model in terms of $a = 2/r$ and $b = f/a$.

[7] https://en.wikipedia.org/wiki/Absorbing_Markov_chai

[8] Aggregation is done using plugins available on other OMSSs and media players.

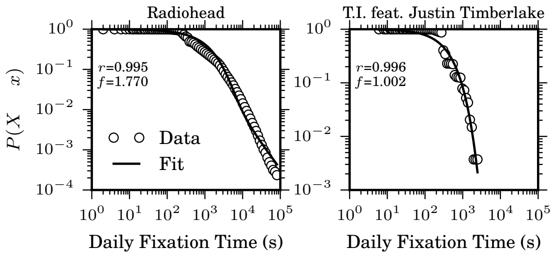

Figure 2. Validation of Fixation Model.

176k plays. We note that even after filtering, there still remains a significant number of rare transitions as 44% of the inter-artist transitions happen less than ten times.

4.1 The Fixation Model at Work

We first discuss the Fixation model. We begin by showing how it fits the time users spend listening to different artists on any given day (referred to as daily fixation time). Figure 2 shows the fitted and empirical complementary cumulative distribution functions (CCDF) of the daily fixation time for two particular example artists, namely *Radiohead*, and *T.I. feat. Justin Timberlake* (a collaboration between two artists). This example was extracted from the Last.FM-2014 dataset.

The distribution for *Radiohead* clearly has long tails, and is similar to the distributions for most artists. In contrast, the distribution for the *T.I. feat. Justin Timberlake* collaboration has a much shorter tail, approaching an exponential distribution. Unlike for the other artists, there is only one song by this artist collaboration in our dataset, which might explain why users tend to spend less time listening to them. Yet, our Fixation model provides close fittings for both distributions, capturing both long and short tails. Interestingly, we can also use the model parameters r and f to distinguish between these artists: compared to *Radiohead*, the *T.I. feat. Justin Timberlake* collaboration has a slightly higher rush parameter ($r = 0.996$) but a much lower fixation parameter ($f = 1.002$). Despite the higher initial surge of attention, users lose interest more quickly in them. If it were not for our separation of the intra-artist from the inter-artist transitions, it would be impossible to capture these different distributions with SWIFT-FLOWS. That is, these superior fits are only possible through the use the modulated Markov models as done by our intra-artist model. This allows the model to capture both long and short tails of user attention [22].

We proceeded to fit our model to the daily fixation times of 36,344 and 2,570 artists in the Last.FM-2014 and Last.FM-2009 datasets respectively (artists with more than 5 plays by at least 5 users). In Figure 3 we show a scatter plot of the fixation versus rush scores for the Last.FM-2014 dataset. We found that, the vast majority of the artists have very high values of rush r (above 0.95) and values of fixation f (1.5 to 2.5). There were also two other small groups of artists with very low (near 1) fixation.

Looking into these groups, we found many collaborations between artists, such as the aforementioned TI feat

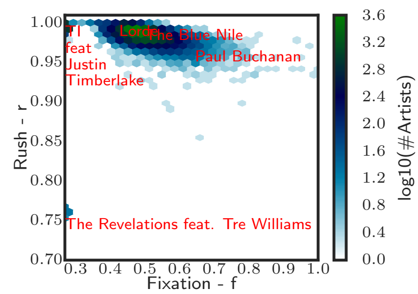

Figure 3. Rush vs Fixation

Justin Timberlake. Another example of a collaboration is The Revelations feat Tre Williams, it has both low fixation and low rush scores, thus attracts little attention in our datasets. We also found that Blue Nile is an interesting example. Blue Nile is a Scottish alternative/pop band whose lead singer is Paul Buchanan. From the plot, we can see that the band has higher rush and lower fixation than the solo songs by Paul Buchanan. This is likely because the solo career of Paul Buchanan mostly attracts more interested fans. Another example is Lorde, a relatively new pop singer at time (2014). The artist obtains high rush and somewhat lower fixation. This may be explained listeners discovering her music.

Our fitting errors are very small in most cases. The average Mean Squared Errors (MSE) of each fitted distribution for artists in Last.FM-2014 is only of 0.02, whereas in the Last.FM-2009 dataset it was of 0.03. The standard deviations were of 0.02 and 0.04 for the Last.FM-2014 and Last.FM-2009 datasets, respectively.

4.2 Extracting Listening Trajectories

We now discuss the Switch model. The first step to execute the model us to decide the number of genes (latent factors) k using the MDL-based criteria described previously. To measure the MDL score, we searched for k in the range $k \in [2, 400]$ [9]. With MDL, we aim at finding a succinct (smaller) yet accurate latent representation of our datasets. In our search, we found that in both datasets as k increases the MDL cost first decreases and then rapidly increases, reaching global minimum at k=40. This value was achieved in both sets of data. For this reason, our experiments use a genome with 40 genes.

Table 1 describes four different genes (latent factors) extracted by SWIFT-FLOWS from the Last.FM-2014 dataset. For each gene, the table shows the top 7 source s and destination d artists in a single column ranked by $(p(a \mid z))$. To further examine the genes, the table also summarizes the nationality and age reported in the LastFM profile of the top 50 users which have attention transitions within each gene. Finally, we cross-referenced the top artists in each gene with the AllMusic guide [10] for an authoritative source on artist metadata. The labels given to each gene stem from our own interpretation.

[9] We searched $k \in \{2, 4, 8, 10, 20, 30, 40, 50, 100, 200, 300, 400\}$.
[10] http://www.allmusic.com/

Table 1. Genes from the Last.FM-2014 dataset: top source and destination artists, and demographics of top-50 users. Artists and users are sorted by probabilities $p(a|z)$ and $p(z|u)$, respectively. Countries are: BR = Brazil, US = USA, NL = Netherlands, DE = Germany, PL = Poland, FI = Finland. Age statistics presented here are the 1^{st}, 2^{nd} and 3^{rd} quartiles. We also show the expected fixation e_z per gene.

	Gene=18 ("BR/US pop")	Gene=20 ("metal")	Gene=23 ("electronic")	Gene=39 ("pop'")
Source/Dest Artists	Britney Spears Wanessa Christina Aguilera t.A.T.u. Katy Perry Pitty Lady Gaga	Nightwish Within Temptation Epica Korn Disturbed Marilyn Manson Rammstein	Daft Punk David Guetta Deadmau5 Skrillex The Prodigy Tiesto Pendulum	Britney Spears Madonna Christina Aguilera Rihanna Lady Gaga Katy Perry Kesha
Users Nationality	BR=98% NL=2%	DE = 18% PL = 16% US = 12% FI = 8%	US = 18% BR = 10% PL = 10% UK = 10%	BR=78% US=10% PL=5%
Age Quartiles	1^{st} = 19 2^{nd} = 21 3^{rd} = 24	1^{st} = 21 2^{nd} = 24 3^{rd} = 29	1^{st} = 20 2^{nd} = 22 3^{rd} = 25	1^{st} = 19 2^{nd} = 22 3^{rd} = 25
e_z	e_z = 793.55	e_z = 642.15	e_z = 636.10	e_z = 886.10

Overall, the genes found through SWIFT-FLOWS point to a semantically sound segmentation of transition spaces that combines characteristics of the artists and users of the OMSS. Illustratively, gene $z = 18$ is predominately formed by female pop/rock singers as both sources and destinations. This is not the only gene with similar pop singers; gene $z = 39$ is another gene with a similar composition in this respect. Yet, the presence of Brazilian pop artists (e.g., *Wanessa*, *Claudia Leitte*, and *Pitty*) in gene $z = 18$ explains why the vast majority (98%) of the top users in this gene are Brazilians (BR). Gene $z = 20$ in turn is mostly focused on different sub-genres of metal (e.g., goth-metal and rap-metal). A large fraction of the top-50 users of the "heavy metal" gene are from Germany and Poland. Finally, gene $z = 23$ represents users of different nationalities (American being the most frequent one) who like to listen to electronic dance music, often transitioning between different artists of that genre. It is also noteworthy that in a dataset mostly comprised of pop artists fans (Last.FM-2014), SWIFT-FLOWS is able to account for the trajectories of heavy metal and electronic music fans.

To understand the expected fixation of users per gene, we make use of Equation 7 (e_z). Initially translate e_z values to seconds. That is, we performed a linear regression using the values of e_a (see Eq. 7), expected number of steps per artist, with the average fixation time per day (described in the previous subsection). With this regression, we found that each step in the chain accounts for, approximately, 1.11 seconds. From the table, we can see that genes 20 and 23 have lower expected fixation times. That is, gene $z = 20$ expects 642 steps (roughly 12 minutes) in the Fixation model, whereas gene $z = 23$ expects 632 steps (11.6 minutes). The highest value in the table is from gene $z = 18$ (14 minutes).

Notice that both models combined provide a general overview of attention. That is, we are able to understand how users will transition between artists, as well as the expected number of steps users will listen to a given artist. This represents one of the major strengths of SWIFT-FLOWS when compared to previous efforts [7, 22].

5. CONCLUSIONS

In this paper, we presented the SWIFT-FLOWS model. One of the main advantages of SWIFT-FLOWS is that it allows researchers to explore user listening habits based on complementary behaviors: the fixation on a single artist over short or long bursts, as well as the change in attention from one artist to the next. We applied SWIFT-FLOWS to uncover semantically meaningful maps of attention flows in large OMSSs datasets. Moreover, SWIFT-FLOWS provides excellent fits to the attention time dedicated to artists. SWIFT-FLOWS, therefore, is an useful tool for further research aiming to understand listening behavior.

Acknowledgments: Research supported by the project FAPEMIG-PRONEX-MASWeb, process number APQ-01400-14, by grant 460154/2014-1 from CNPq, by the EU-BR BigSea project (MCTI/RNP 3rd Coordinated Call), and by the EU-BrazilCC (grants FP7 614048 and CNPq 490115/2013-6) project, as well as by individual grants from CNPq/CAPES/Fapemig. Our work was also supported by NSF grant CNS-1065133, U.S. Army Research Laboratory under Cooperative Agreement W911NF-09-2-0053, and by the Defense Threat Reduction Agency contract No. HDTRA1-10-1-0120. The views and conclusions contained here are our own (the authors).

6. REFERENCES

[1] Alejandro Bellogín, Arjen P. de Vries, and Jiyin He. Artist Popularity: Do Web and Social Music Services Agree? In *Proc. ICWSM*, 2013.

[2] Oscar Celma. *Music Recommendation and Discovery in the Long Tail*. Springer, 1 edition, 2010.

[3] Shuo Chen, Josh L Moore, Douglas Turnbull, and Thorsten Joachims. Playlist prediction via metric embedding. In *Proc. KDD*, 2012.

[4] Shuo Chen, Jiexun Xu, and Thorsten Joachims. Multi-space probabilistic sequence modeling. In *Proc. KDD*, 2013.

[5] Imre Csiszár and Paul C. Shields. The consistency of the bic markov order estimator. *The Annals of Statistics*, 28(6):1601–1619, 12 2000.

[6] Flavio Figueiredo, Jussara M Almeida, Yasuko Matsubara, Bruno Ribeiro, and Christos Faloutsos. Revisit Behavior in Social Media: The Phoenix-R Model and Discoveries. In *PKDD/ECML*, 2014.

[7] Flavio Figueiredo, Bruno Ribeiro, Jussara M Almeida, and Christos Faloutsos. TribeFlow: Mining & Predicting User Trajectories. In *Proc. WWW*, 2016.

[8] Tom Griffiths. Gibbs sampling in the generative model of latent dirichlet allocation. Technical report, 2002.

[9] Mark H Hansen and Bin Yu. Model Selection and the Principle of Minimum Description Length, 2001.

[10] David J Hargreaves. Musical imagination: perception and production, beauty and creativity. *Psychology of music*, 40(5):539–557, 2012.

[11] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times*. AMS, 2008.

[12] Andryw Marques, Nazareno Andrade, and Leandro Balby Marinho. Exploring the Relation Between Novelty Aspects and Preferences in Music Listening. In *Proc. ISMIR*, 2013.

[13] Yasuko Matsubara, Yasushi Sakurai, Christos Faloutsos, Tomoharu Iwata, and Masatoshi Yoshikawa. Fast mining and forecasting of complex time-stamped events. In *Proc. KDD*, 2012.

[14] Carl D. Meyer. Stochastic Complementation, Uncoupling Markov Chains, and the Theory of Nearly Reducible Systems. *SIAM Review*, 31(2):240–272, 1989.

[15] Joshua L. Moore, Shuo Chen, Thorsten Joachims, and Douglas Turnbull. Taste Over Time: The Temporal Dynamics of User Preferences. In *Proc. ISMIR*, 2013.

[16] Raphaël Nowak. Investigating the interactions between individuals and music technologies within contemporary modes of music consumption. *First Monday*, 19(10):Online, 2014.

[17] Chan Ho Park and Minsuk Kahng. Temporal Dynamics in Music Listening Behavior: A Case Study of Online Music Service. In *Proc. ACIS*, 2010.

[18] Minsu Park, Ingmar Weber, Mor Naaman, and Sarah Vieweg. Understanding Musical Diversity via Online Social Media. In *Proc. ICWSM*, 2015.

[19] Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. Factorizing personalized Markov chains for next-basket recommendation. *Proc. WWW*, 2010.

[20] P. J. Rentfrow and S. D. Gosling. The Do Re Mi's of Everyday Life: The Structure and Personality Correlates of Music Preferences. *Journal of personality and social psychology*, 84(6), 2003.

[21] Peter J. Rentfrow, Lewis R. Goldberg, and Daniel J. Levitin. The structure of musical preferences: a five-factor model. *Journal of personality and social psychology*, 100(6):1139–1157, June 2011.

[22] Stephan Robert and Jean-Yves Le Boudec. On a Markov modulated chain exhibiting self-similarities over finite timescale. *Performance Evaluation*, 27-28:159–173, 1996.

[23] Suvi Saarikallio and Jaakko Erkkilä. The role of music in adolescents' mood regulation. *Psychology of music*, 35(1):88–109, 2007.

[24] Philipp Singer, Denis Helic, Andreas Hotho, and Markus Strohmaier. HypTrails: a bayesian approach for comparing hypotheses about human trails on the web. In *Proc. WWW*, 2015.

[25] Liang Xiong, Xi Chen, Tzu-Kuo Huang, Jeff Schneider, and Jaime G. Carbonel. Temporal collaborative filtering with bayesian probabilistic tensor factorization. In *Proc. SDM*, 2010.