Inferring the Causal Impact of New Track Releases on Music Recommendation Platforms through Counterfactual Predictions

Rishabh Mehrotra
Spotify
rishabhm@spotify.com

Prasanta Bhattacharya
Institute of High Performance Computing (IHPC), A*STAR
prasanta_bhattacharya@ihpc.a-star.edu.sg

Mounia Lalmas
Spotify
mounial@spotify.com

ABSTRACT

With over 20,000 tracks being released each day, recommendation systems that power music streaming services should not only be responsive to such large volumes of content, but also be adept at understanding the impact of such new releases on, *both*, users' listening behavior and popularity of artists. Inferring the causal impact of new track releases is critical to fully characterizing the interplay between artists and listeners, as well as among the artists. In this study, we infer and quantify causality using a diffusion-regression state-space model that constructs counterfactual outcomes using a set of synthetic controls, which predict potential outcomes in absence of the intervention. Based on large scale experiments spanning over 21 million users and 1 billion streams on a real world streaming platform, our findings suggest that releasing a new track has a positive impact on the popularity of other tracks by the same artist. Interestingly, other related and competing artists also benefit from a new track release, which hints at the presence of a positive *platform-effect* wherein some artists gain significantly from activities of other artists.

CCS CONCEPTS

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

KEYWORDS

Causal Impact, Recommender Systems, Counterfactual Predictions

ACM Reference Format:
Rishabh Mehrotra, Prasanta Bhattacharya, and Mounia Lalmas. 2020. Inferring the Causal Impact of New Track Releases on Music Recommendation Platforms through Counterfactual Predictions. In *Fourteenth ACM Conference on Recommender Systems (RecSys '20), September 22–26, 2020, Virtual Event, Brazil.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3383313.3418491

1 INTRODUCTION

The music content generated on streaming platforms such as Spotify or Pandora generally come from a wide variety of artists whose tracks, once launched, are listed on the streaming service and is gradually diffused to a large number of listeners. Thus, the launch of a new music track not only increases the visibility and discovery of the artist whose track is released, but also boosts overall engagement for the platform itself. For instance, whenever Akon, an American singer, launches a new track, it attracts some first time Akon-listeners to discover and search for his songs, which increases the artist's popularity. Indirectly, listeners often end up spending more time on the music platform, thereby increasing the chances of discovering other similar artists, and improving their overall engagement with the platform. It is therefore increasingly critical that music recommendation systems on these platforms take into account, and are prepared to deal with the potential changes in consumption preferences and platform engagement that accompany such track releases. With over 10,000 songs released per day globally[1], inferring the causal impact of a new track release on listener behavior is likely to become an important consideration in the design of such recommendation systems, as well as other demand forecast tools for artists.

Inferring causal impact, as opposed to merely analyzing significant correlations, is important as causal inference models help shed light on the underlying mechanisms that potentially drive these correlations. However, estimating the impact of a new track release on overall listener engagement with the focal artist, and other similar artists is a challenging empirical problem. It is not trivial to derive unbiased causal estimates in this context due to practical limitations in running randomized experiments, absence of counterfactual data, and limitations with obtaining fine-grained usage data. In our study, we address these limitations by leveraging a large-scale music activity dataset, and a quasi-experimental Bayesian framework that quantifies the effect of a treatment (i.e. track release) on an outcome of interest - artist stream counts, in this case. The causal impact of the intervention can then be estimated as the difference between the observed outcome and the unobserved or potential outcome, had the users not been treated (i.e. not exposed to a new track release).

We adopt a Bayesian and structural time-series modeling strategy to assess the causal impact of a new track release on the popularity of the focal artist (i.e. artist who released a track) as well as other similar artists. The proposed diffusion-regression state-space model constructs a counterfactual artist popularity outcome using a set of synthetic controls, that predict the potential outcome in absence of any intervention. Based on the analysis of music streaming data from over 21 million users, we investigate the impact of a new track launch on (i) the engagement for the focal artist, (ii) the engagement of similar artists, and (iii) the interplay between the engagements for the focal and similar artists. We show that new track releases significantly improve listener engagement for the focal artist. We also find robust evidence for a *platform-effect*, wherein an increased

RecSys '20, September 22–26, 2020, Virtual Event, Brazil

ACM ISBN 978-1-4503-7583-2/20/09.
https://doi.org/10.1145/3383313.3418491

[1] http://www.nielsen.com/us/en/press-room/2017/nielsen-releases-2016-us-year-end-music-report.html

popularity of the focal artist results in an increased engagement for other similar artists. We discuss future research directions stemming from the findings presented in this work.

2 RELATED WORK

Prior research on music understanding and consumption has focused on automated playlist generation from user listening patterns [4], role of personality traits in predicting music taxonomy preferences [9], understanding users' music listening needs and intents [9, 14], user's consumption diversity of music [1], the interplay between user and artist objectives [15, 16] and the impact of adoption of online streaming on music consumption [8]. The studies in this area help us hypothesize about the various factors that might influence a listener's decision to engage with a particular artist and track.

Forecasting trends in access to information, as well as modeling and predicting popularity in user generated content has important applications in many emerging areas, such as support facilitation, advertising, content caching, price estimations, traffic management, as well as financial and economic forecasting. Prior work has exploited correlation in content views over time [6, 18], survival analysis [11, 12], linear regression models [2, 3, 20], entropy measure based on the "user-interest peak" and the "co-participatory network" [10] to predict content popularity.

Advancements in causal inference modeling have led to the emergence of new methods for causal structure learning and predictions (i.e., predicting the aftermath of interventions). In prior studies, causal relationships among related time series have been modeled using Granger causality approaches [19], lagged correlation [13], and Bayesian networks [22], among others. A variety of causality mining techniques have been studied in past work for content popularity in social media, including Granger causality based influence model for Twitter [7], Granger graphical models for anomaly detection [17] and state-space diffusion regression models to predict counterfactual market response [5], which forms the basis of the current method.

3 ESTIMATING CAUSAL IMPACT

Each new track release brings user traffic to music streaming platforms. This increased user engagement often results in users exploring music by the same artist and other similar artists. Next, we describe the proposed model used to infer such causal estimates. Here, **track** is a song or recorded music by an artist; **focal artist** is the main artist who has released a new track; and **similar artists** are artists similar to the focal artist in terms of their genre-profiles.

3.1 Causal Impact Model

To infer the causal impact of a track release on the music consumption attributable to an artist, we model the music consumption behavior over time as a structural state-space model [5]. A state-space model specifies the various states of the model and the associated representations. State-space models are particularly useful in practice on account of their flexibility and modularity. The latent spaces comprising the state-space model are essentially composed of a library of sub-models that help to model the important attributes of the context. In our case, this flexibility enables us to capture trend, seasonality or other domain specific effects. We define this state-space model as follows:

$$y_t = Z_t^T \alpha_t + \epsilon_t \tag{1}$$

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t \tag{2}$$

The first equation is the observation equation, and models the observed scalar variable y_t as a function of a latent d-dimensional state vector α_t. The second equation specifies the evolution of this state vector α_t over time. $\epsilon_t \sim N(0, \sigma_t^2)$ and $\eta_t \sim N(0, Q_t)$ are independent of all other unknowns. In our track release problem setting, y_t specifies the number of streams from a given artist on a given day, Z_t is a d-dimensional output vector, T_t is a d x d transition matrix, and R_t is a d x q control matrix. Any observation error is captured by ϵ_t, while η_t is a q-dimensional system error with a q x q state-diffusion matrix Q_t , where $q \leq d$. We assume the error terms from these state-component models to be independent. α_t can subsequently be estimated by concatenating the state components. The most important state component for deriving causal estimates is the regression component, which leverages music consumption that was not treated i.e., not impacted by the new track release, to create an artificial or synthetic control group, which in turn serves as a quasi-counterfactual in the estimation of the treatment effect. We next describe the different state components that we specify in the causal inference model.

3.2 State Components

Prior research on music understanding and consumption has focused on automated playlist generation [4], role of personality traits in predicting music taxonomy preferences [9], understanding users' music listening needs [9] and the impact of adoption of online streaming on music consumption [8]. These studies point to two prominent factors that might influence a listener's decision to engage with a particular artist and track - local trends and seasonality. We model such factors as different state components.

Local Trend. Often the popularity of an artist follows a local trend, owing to a sudden increased interest in the artist among listeners. The local linear trend component of our model, as shown below, allows us to model changes attributable to any such local shocks to the context. This enables us to make accurate short-term and local predictions.

$$\mu_{t+1} = \mu_t + \eta_{\mu,t} \tag{3}$$

$$\delta_{t+1} = \delta_t + \eta_{\delta,t} \tag{4}$$

where $\eta_{\mu,t} \sim N(0, \sigma_\mu^2)$ and $\eta_{\delta,t} \sim N(0, \sigma_\delta^2)$. The μ_t component captures the trend at time t, while the δ_t captures the expected increase in μ from t to $t + 1$.

Seasonality. Music listening patterns often follow seasonal trends, like day-of-the-week trend, or weekend trend, e.g. [21]. To allow the model to incorporate such seasonal variations, we include a seasonality component in our state component as follows:

$$\lambda_{t+1} = -\sum_{s=0}^{S-2} \lambda_{t-s} + \eta_{\lambda,t} \tag{5}$$

where S and λ_t capture the number of periods as well as their joint contributions to the observed outcome y_t, respectively.

Contemporaneous Covariates. We use a set of control time series that received no treatment for making accurate counterfactual predictions since they account for any shared variance across the predictor time-series, and help to control for the effects of any unobserved sources of variance. We incorporate such control series in the model using a simple linear regression, whose variables can be time-varying or time-invariant.

3.3 Spike-and-Slab Prior for Synthetic Control

The counterfactual is constructed by concatenating a set of selected predictor series into a single synthetic control [5]. The estimator then selects from a collection of such potential controls, by applying a spike-and-slab prior on the set of regression coefficients. The estimator then averages over this set of controls. The use of a spike-and-slab as a prior is important as it combines the point mass at zero (hence, the "spike"), for a certain subset of zero coefficients, with a weakly informative distribution on a complementary set of non-zero coefficients (hence, the “slab").

Let $\rho = (\rho_1, ..., \rho_J)$, where j = 1 if $\beta_j = 0$, and 0 otherwise. Also, let β_ρ denote the non-zero elements of β and $\sum_\rho^{-1}$ denote those entries of $\sum^{-1}$ that correspond to the non-zero entries in ρ. We then factorize the prior as:

$$p(\rho, \beta, \frac{1}{\sigma_\epsilon^2} = p(\rho)p(\sigma_\epsilon^2|\rho)p(\beta_\rho|\rho, \sigma_\epsilon^2) \quad (6)$$

In the above, the spike is, in principle, an arbitrary distribution over $(0, 1)^J$. A possible and common choice here is a simple product of Bernoulli distributions. For the “slab," it is common to use a conjugate normal-inverse Gamma distribution.

3.4 Estimation of Causal Impact

We draw on [5] to adapt a Bayesian model of causal inference. We specify a prior distribution $p(\theta)$, and a distribution $p(\alpha_0|\theta)$ on the initial state values. Next, we perform a MCMC sampling of $p(\alpha, \theta|y)$. While most similar Bayesian models make inference based on the posterior distribution over parameters and states $p(\theta, \alpha|y_{1:n})$, we infer the causal estimates following the posterior incremental effect as shown below:

$$p(y^*_{n+1:m}|y_{1:n}, X_{1:m}) \quad (7)$$

As is evident, the density above is defined for the span of time period for which the outcomes are unobserved. This serves to capture the true counterfactual i.e. count of artist stream that would be observed in the treated artists' streams following the intervention, but in the absence of any intervention.

We then sample from the posterior distribution over counterfactual activity to generate a statistic of the posterior causal effect, i.e. our main estimate of interest. For each draw τ and time point $t = n + 1, ..., m$, we set

$$\phi_t(\tau) := y_t - y_t^{*(\tau)} \quad (8)$$

This samples from the approximate posterior predictive density of the intervention. In addition to the point-wise impact, it is important to compute a cumulative impact estimate of this intervention over time, which can be computed as:

$$\frac{1}{t-n} \sum_{t^1=n+1}^{t} \phi_{t^1}^{(\tau)} \forall t = n+1, ..., m \quad (9)$$

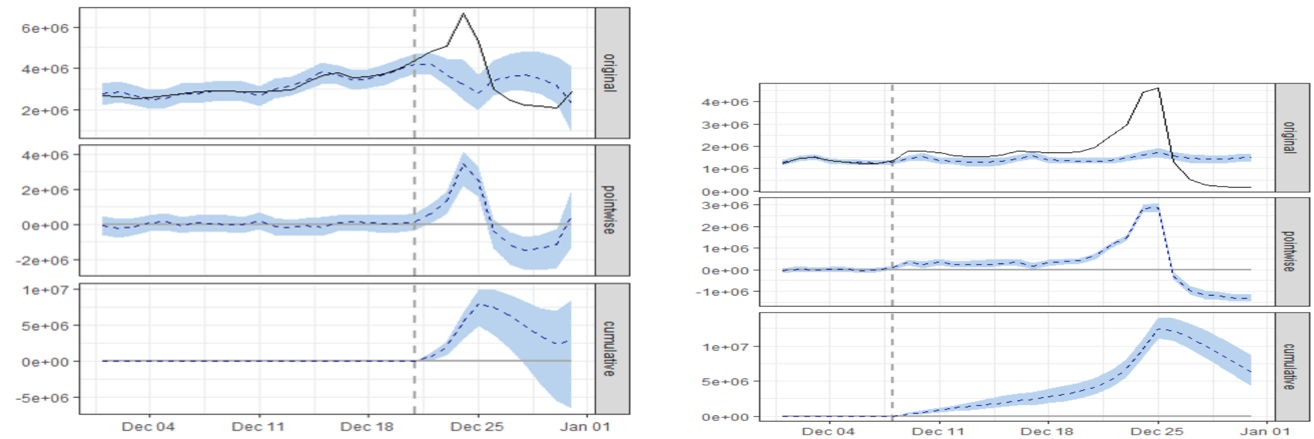
In the next section, we report estimation results using a large and real-world dataset of track release and music consumption behavior over time.

4 EXPERIMENTS

The Bayesian structural model described in the previous section allows us to estimate the causal impact of a new track release on the popularity of the focal and related artists. To derive such causal estimates, and as mentioned earlier, we leverage a large scale music consumption dataset from a commercial music streaming service, and consider all new tracks released in the month of December 2018 by the top 5000 most popular artists. In addition, we monitor music consumption behavior of a sample of 21 million users and over 1.4 billion streams. We apply our proposed model to the track and artist stream behavior before and after the release of a new track, with the per-day stream count as the outcome variable.

Inferring Causal Impact of Track Release. Figure 1(left) illustrate the distribution of point estimates for the relative intervention effect, and the cumulative post-intervention effect, for a highly popular focal artist. The first panel on the left contrasts the observed stream counts (solid line) against the counterfactual prediction of the stream count (dashed line) in the period following the intervention. The counterfactual prediction (dashed-line) indicates the stream count had the new track not been released. The difference between these two indicates that a new track release resulted in increased stream counts for this artist. The second panel (Figure 1(Left)) plots this point-wise causal estimate of the intervention, defined as the difference between observed stream count and the counterfactual in the post-intervention period. We observe that the track release caused an increase in artist's music stream count, which subsided after a few days. The third panel (Figure 1 left) presents a cumulative effect of the intervention, obtained by adding up the point-wise estimates in the period following the intervention. We observe that the new track release caused a steady and cumulative increase in the total stream counts. Figure 1 (right) shows the point-wise causal estimate and the cumulative causal effect of the intervention for a *less popular* artist. We observe that while this artist also experienced a surge in stream count following the release of a new track, there is a significant delay before it reached peak popularity.

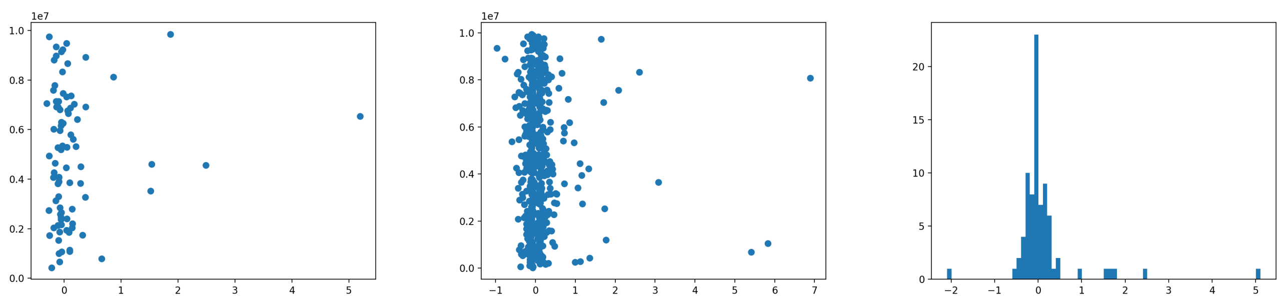
Quantifying Impact on Streaming Popularity. We applied the same causal effect model to ~300 artists, who had at least one track launch within our study period. We estimated our model to quantify the gain/loss in streaming popularity following the new track release. Figure 2 (left) denotes the point estimates of the causal effect at the 10% level of significance. We can find artists with both, a statistically significant increase and a decrease in stream counts in the post-intervention period after subtracting the corresponding stream counts in the counterfactual time series. However, while most of the negative estimates are clearly clustered around 0, positive estimates have a larger variance with some exceeding 100%. This shows that releasing a new track contributes positively to the stream count for an artist by drawing attention of the listeners.



(a) Effect on Focal Artist

(b) Effect on Less Popular Artist

Figure 1: Heterogeneity in treatment effect of new track release. Left: Estimating the treatment effect for the focal artist. Right: Estimating the treatment effect for a less popular artist. For both figures, the first panel plots the data together with the counterfactual predictions following the track release. The second panel plots the point-wise treatment effect i.e. the difference between observed data the and counterfactual predictions. The third panel plots the cumulative effect of the point-wise treatment effect over time in the post-release period.



(a) Relative impact on Focus Artists

(b) Relative impact on Related Artists

(c) Relative Differences

Figure 2: Relative impact of track launch on focal artist (Left) and related artist (Middle). X-axis represents estimated causal effect, with a value of 1 denoting a 100% increase. Right: Relative difference in the causal impact of a new track release on focal vs related artists.

Causal Impact on Related Artists. We next investigate if track launches tend to affect listening rates of similar artists. To study this spill-over effect on the streaming counts of other similar artists, we used a measure of artist similarity based on their music genre-profiles. For each artist, we picked the top similar artists, and studied the time series perturbations of stream counts of these similar artists in the post-intervention period. Figure 2 (middle) illustrates the distribution of point estimates of the causal effect for the total set of ~3100 similar artists at 10% level of significance. We can see that the launch of the music track introduces both positive and negative externalities for the related artists. However, as evident from Figure 2, we find a significant and positive change for most of the related artists, with some related artists experiencing an increase of over 100% in their stream counts.

Relative Differences between Focal & Related Artists. Our results show that a track release can positively affect artist popularity, and also lead to positive externality for other similar artists. However, it is also important to understand how these benefits for the focal artist compare against the benefits to the artist's similar peers. Figure 2 (right) illustrates the distribution of the difference between the causal effects of 87 focal artists (i.e. a set of focal artists with statistically significant casual effect at 10% level of significance) and the average of causal effects of their peers. We find that, for most artists, the benefits are equally shared by the focal artist and their peers. However, we also find evidence of focal artists benefiting more than their peers, as well as of peers benefiting more than the focal artist.

5 DISCUSSION

Our findings indicate that new track releases are an important event not just for the focal artist but also for other artists, who might benefit by virtue of being on the streaming platform. This is probably due to the increased tendency of music lovers to explore and discover other albums and artists related to the newly released track. It was also insightful to note that artists with varying platform popularity exhibit different trajectories in stream count growth following the release of major tracks, with less popular artists taking longer to reach peak stream counts. We contend that music recommendation systems need to be better aware of such heterogeneities in consumption patterns following major events on the platform, to allow less popular or niche artists to also benefit from the event.

Most music streaming platforms have recommendation systems that help in facilitating the music discovery process. The findings from this paper highlight the need to look at ways to increase salience of focal artists to surface them to users and consequently improve the discovery and engagement of related genres and artists. Specific future directions for this study include: (i) developing models to estimate and improve user retention on streaming platforms by leveraging such release events, (ii) designing new ways of improving visibility of popular artists, while also promoting the right set of related genres and artists.

As mentioned above, we also shed light in this paper on consumer behavior following popular track releases. By observing the perturbations in stream counts following the launch, we can begin to formulate ideas about how audiences react to new music, in terms of engagement, discovery and overall satisfaction. Our analyses show that the audience reactions to new album launches might also be negative, and there are often various types of artist-level heterogeneities that influence how audiences react to such events. Furthermore, some users might get engaged with the platform only to interact with such new track releases. Music recommendation systems of the future need to be aware of these contextual dependencies. Future work can leverage the findings and insights from this study to inform track release strategies for artists, while improving recommendations for listeners.

REFERENCES

[1] 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*. 2155–2165.

[2] Ioannis Arapakis, Berkant Barla Cambazoglu, and Mounia Lalmas. 2017. On the feasibility of predicting popular news at cold start. *JASIST* 68, 5 (2017), 1149–1164. https://doi.org/10.1002/asi.23756

[3] Roja Bandari, Sitaram Asur, and Bernardo A Huberman. 2012. The Pulse of News in Social Media: Forecasting Popularity.. In *ICWSM*. 26–33.

[4] Shay Ben-Elazar, Gal Lavee, Noam Koenigstein, Oren Barkan, Hilik Berezin, Ulrich Paquet, and Tal Zaccai. 2017. Groove Radio: A Bayesian Hierarchical Model for Personalized Playlist Generation. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM, 445–453.

[5] Kay H Brodersen, Fabian Gallusser, Jim Koehler, Nicolas Remy, Steven L Scott, et al. 2015. Inferring causal impact using Bayesian structural time-series models. *The Annals of Applied Statistics* 9, 1 (2015), 247–274.

[6] Meeyoung Cha, Alan Mislove, and Krishna P Gummadi. 2009. A measurement-driven analysis of information propagation in the flickr social network. In *Proceedings of the 18th international conference on World wide web*. ACM, 721–730.

[7] Yi Chang, Xuanhui Wang, Qiaozhu Mei, and Yan Liu. 2013. Towards Twitter context summarization with user influence models. In *Proceedings of the sixth ACM international conference on Web search and data mining*. ACM, 527–536.

[8] Hannes Datta, George Knox, and Bart J Bronnenberg. 2017. Changing their tune: How consumers' adoption of online streaming affects music consumption and discovery. *Marketing Science* (2017).

[9] Bruce Ferwerda, Emily Yang, Markus Schedl, and Marko Tkalcic. 2015. Personality traits predict music taxonomy preferences. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2241–2246.

[10] Salman Jamali and Huzefa Rangwala. 2009. Digging digg: Comment mining, popularity prediction, and social network analysis. In *Web Information Systems and Mining, 2009. WISM 2009. International Conference on*. IEEE, 32–38.

[11] Jong Gun Lee, Sue Moon, and Kavé Salamatian. 2010. An approach to model and predict the popularity of online contents with explanatory factors. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, Vol. 1. IEEE, 623–630.

[12] Jong Gun Lee, Sue Moon, and Kavé Salamatian. 2012. Modeling and predicting the popularity of online contents with Cox proportional hazard regression model. *Neurocomputing* 76, 1 (2012), 134–145.

[13] Ajay Anil Mahimkar, Zihui Ge, Aman Shaikh, Jia Wang, Jennifer Yates, Yin Zhang, and Qi Zhao. 2009. Towards automated performance diagnosis in a large IPTV network. In *ACM SIGCOMM Computer Communication Review*, Vol. 39. ACM, 231–242.

[14] Rishabh Mehrotra, Mounia Lalmas, Doug Kenney, Thomas Lim-Meng, and Golli Hashemian. 2019. Jointly leveraging intent and interaction signals to predict user satisfaction with slate recommendations. In *The World Wide Web Conference*. 1256–1267.

[15] Rishabh Mehrotra, James McInerney, Hugues Bouchard, Mounia Lalmas, and Fernando Diaz. 2018. Towards a fair marketplace: Counterfactual evaluation of the trade-off between relevance, fairness & satisfaction in recommendation systems. In *Proceedings of the 27th acm international conference on information and knowledge management*. 2243–2251.

[16] Rishabh Mehrotra, Niannan Xue, and Mounia Lalmas. 2020. Bandit based Optimization of Multiple Objectives on a Music Streaming Platform. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 3224–3233.

[17] Huida Qiu, Yan Liu, Niranjan Subrahmanya, Weichang Li, et al. 2012. Granger causality for time-series anomaly detection. In *Data Mining (ICDM), 2012 IEEE 12th International Conference on*. IEEE, 1074–1079.

[18] Gabor Szabo and Bernardo A Huberman. 2010. Predicting the popularity of online content. *Commun. ACM* 53, 8 (2010), 80–88.

[19] Jin Tian and Judea Pearl. 2000. Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence* 28, 1-4 (2000), 287–313.

[20] Manos Tsagkias, Wouter Weerkamp, and Maarten De Rijke. 2010. News comments: Exploring, modeling, and online prediction. In *Advances in Information Retrieval*. Springer, 191–203.

[21] Sergey Volokhin and Eugene Agichtein. 2018. Understanding Music Listening Intents During Daily Activities with Implications for Contextual Music Recommendation. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval* (New Brunswick, NJ, USA) *(CHIIR '18)*. ACM, New York, NY, USA, 313–316. https://doi.org/10.1145/3176349.3176885

[22] Changhe Yuan, Xiaolu Liu, Tsai-Ching Lu, and Heejin Lim. 2009. Most Relevant Explanation: properties, algorithms, and evaluations. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*. AUAI Press, 631–638.