Social Networks 61 (2020) 144–152

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Social complex contagion in music listenership: A natural experiment with 1.3 million participants

John Ternovski[a,b,c], Taha Yasseri[c,d,*]

[a] Yale University, New Haven, CT 06520-8209, USA
[b] Harvard Kennedy School, Harvard University, Cambridge, MA 02138, USA
[c] Oxford Internet Institute, University of Oxford, 1 St Giles, Oxford OX1 3SJ, UK
[d] Alan Turing Institute, 96 Euston Rd, London NW1 2DB, UK

ARTICLE INFO

Keywords:
Contagion
Influence
Social networks
Indirect effects
Music

ABSTRACT

Can live music events generate complex contagion in music streaming? This paper finds evidence in the affirmative—but only for the most popular artists. We generate a novel dataset from a music tracking website to analyse the listenership history of 1.3 million users over a two-month time horizon. We show that attending a music artist's live concert increases that artist's listenership among the attendees of the concert by approximately 1 song per day per attendee (p-value < 0.001). Moreover, this effect is contagious and can spread to users who did not attend the event. However, whether or not contagion occurs depends on the type of artist. We only observe contagious increases in listenership for popular artists (~ 0.06 more daily plays per friend of an attendee [$p < 0.001$]), while the effect is absent for emerging stars. The contagion effect size increases monotonically with the number of friends who have attended the live event.

Introduction

David Bowie has been quoted as saying "[m]usic itself is going to become like running water or electricity, [so] you'd better be prepared for doing a lot of touring" (Krueger, 2005). Bowie's prescient prediction of a post-Napster world led one economist to coin the "Bowie Theory" to explain the changing economic model used by the music industry (Krueger, 2005). Namely, the Bowie Theory summarizes the industry's shift from relying on revenue from physical copies of pre-recorded music to that of live performances. While recorded sales have fallen, revenues from live performances have held steady (Charron, 2017; Montoro-Pons and Cuadrado-García, 2011). But does this hold only for the most popular bands? After all, music insiders claim that even "mid-level" bands would "be doing well to break even" touring (Passman, 2006). Altruistic motivations aside, why tour? Are there *any* economic benefits to live performances? Research on the music market has found that besides the direct benefits of touring (i.e. ticket and merchandise sales), tours provide an opportunity for an artist "to expand their fan base" (Black et al., 2007). This latter "indirect effect" has generally been rather amorphous in the literature, but advances in network analysis allows us to frame this question in terms of contagion. In other words, a fan who attends a live event may have some measurable impact on their friends' music listening habits—that impact can be seen as an "infection." The question then is, holding constant the occurrence of live events, how can this infection spread over social networks? Or put differently, how many of your friends must attend a live event for you to change your music listenership habits as if you attended the event yourself?

Social contagion has been studied in different systems and under different dispersion scenarios; these include political mobilization through peer networks (Bond et al., 2012; Jones et al., 2017), adoption of health behaviours among members of online communities (Centola, 2010) and real-world social networks (Aral and Nicolaides, 2017), leveraging peers for viral marketing (Leskovec et al., 2007), and the spread of hashtags on Twitter (Romero et al., 2011; Mønsted et al., 2017). Social influence has also been found to have a critical role in the art market (Salganik and Watts, 2008). But are the social contagion processes more effective for a certain category of artists? Social influence signals are widely used in such settings and help promote popular products to maximize market efficiency. However, it has been argued that social influence makes these markets unpredictable (Salganik et al., 2006). As a result, social influence has been presented in a negative light. However, when it comes to market activities, such processes can lead to considerable revenues.

The music industry in just the US has been valued at $17.2 billion as of 2016 (Resnikoff, 2017), so there has been no shortage of incentives

* Corresponding author at: Oxford Internet Institute, University of Oxford, 1 St Giles, Oxford OX1 3SJ, UK.
E-mail addresses: john.ternovski@yale.edu (J. Ternovski), taha.yasseri@oii.ox.ac.uk (T. Yasseri).

https://doi.org/10.1016/j.socnet.2019.10.005

Available online 31 October 2019

to optimize the revenue streams within it. However, it is the advent of illegal sharing of pre-recorded music by means of mp3s in the early 2000s that inspired renewed interest in developing new economic models for music consumption (Liebowitz, 2003). In the modern music industry, there have been three major streams of revenue identified: (1) the Internet and digital music consumption, (2) CDs, records and conventional pre-recorded music consumption, and (3) live music consumption (Rogers, 2013). The new economic models that emerged gave greater weight to live performances (Dewenter et al., 2012) and the industry reacted. As of 2016, live music revenues accounted for more than half of all US music revenue (Washenko, 2017). Prior to the 2000s, there have been relatively few studies that sought to understand the mechanics of live events (Black et al., 2007). This is not surprising as the conventional wisdom in the music industry before the Internet was that live performances should be treated as nothing more than promotions for pre-recorded music (Montoro-Pons and Cuadrado-García, 2011).

Scholars identified the major reasons why artists choose to perform live as some combination of the following: (1) direct profits, (2) expanding listenership, and (3) strengthening their existing fan base (Black et al., 2007). Prior to the Internet, the industry thought live performances could only satisfy the latter two factors (Passman, 2006), but, today, studies have clearly demonstrated that live music can capture a large share of direct revenues. Black et al. (2007) find a trend of increasing ticket prices and unflagging demand, while industry reports illustrate live event revenues have continued to grow (Washenko, 2017).

The question that remains is whether the traditional reason for concerts—namely, “promotional” effects—still exist. Some researchers argue that there is something unique about live performances, as attendance has not waned despite the substantially higher costs (Earl, 2001). If the indirect promotional effects ever existed, it is likely that they still exist. Indeed, the increased ability to communicate information by the average concert-goer to a large number of personal contacts through online social networks may mean that the indirect effects are larger.

Sociologists have claimed that the social component is essential to “musicking” (Small, 1998) and much of the research in the sociology of music has explored musical social networks (e.g., Crossley et al., 2014; Crossley and Bottero, 2015a). Mark (1998) wrote that such networks are characterized by high levels of homophily and are the vehicle by which music preferences spread from one person to another. Some sociological (e.g., Mark, 2003) and economic (e.g. Dewenter et al., 2012) models predict indirect effects, but actually connecting music consumption to offline behaviour has proven challenging due to sparse data availability. Montoro-Pons and Cuadrado-García, 2011 made significant in-roads evaluating the link between concert attendance and music consumption, concluding that concert attendance does not cause increases in CD purchases. They do note that their analysis does not capture other modes of music consumption and concert attendance may still have impacts on music listenership (Montoro-Pons and Cuadrado-García, 2011). Another study linked pre-recorded music listenership with offline concert attendance to evaluate if fan preferences concord with songs actually played live (Rodriguez et al., 2008). Most recently, Maasø (2016) looked at streaming patterns before and after a music festival in Norway. However, that study was a macro-level analysis and so it did not attempt to isolate the peer effects from other possible confounds or attempt a more stringent causal identification strategy (unlike the design used in this paper).

Because real online social networks are densely interconnected, causal identification in a network context is difficult (Rogowski and Sinclair, 2012). Namely, there is a copious amount of interference between units which violates the SUTVA assumption necessary for implementing—in an ideal scenario—a randomized controlled experiment. Using simple lagged variables in a logistic regression was initially a very popular, simple approach to measuring peer effects, however this method was prone to large biases (Cohen-Cole and Fletcher, 2008a, b). As such, network researchers today either use complicated random assignment protocols (Basse and Airoldi, 2018) or localize treatment effects in a way that incorporates interference (Aral and Nicolaides, 2017). But even in many contemporary designs, the researcher must exclude large swaths of data to preserve SUTVA. What's worse, the data to be excluded depends on whether one is estimating the direct or the indirect effect.

This is where time-series data and a regression discontinuity design can be helpful. If the treatment effects of a discontinuous shock are both highly localized (no spillover beyond the immediate neighbours) and short-term (no long-lasting effects that interact with future treatment effects), one is able to compare the effect of treatment *within units* for direct effects (and indirect effects) without excluding inordinately large amounts of data. The method used in this paper is rooted firmly in the estimators explicated by Hahn et al. (2001) and, in more detail, by Imbens and Lemieux (2008). Our application of a regression discontinuity design to time-series data follows Malik and Pfeffer (2016). Substantively, we build on the approach used by Rodriguez et al. (2008), which leveraged Last.fm data to link listenership habits to live event attendance. We evaluate if live events have any impact on music consumption. More crucially, we investigate whether live event attendance has any *indirect* effects on listenership among the attendee's friends. We divide non-pecuniary benefits into either direct effects or indirect effects. Direct effects increase a given individual's music consumption of that band *as a result of* their attendance. These effects include the expansion of the band's fan base (i.e., non-fans who go to a live event and become fans, consuming the artist's music after the event) and satisfying existing fans (i.e., fans who see the live event and subsequently consume the artist's music at a greater rate). Indirect effects include all gains in music consumption from the live event by individuals *who did not themselves attend the event.* These effects capture the recommendations from event attendees and other forms of communication about the events.

To evaluate indirect effects, we extract all of the attendees' Last.fm friends (for a total of $\sim$1.3 million users) and their listenership data. Any discontinuity in listenership at the time of the live event (that their friend(s) went to) can be interpreted as the indirect impact of event attendance on non-attendees. This measure captures indirect effects that result from either passive signalling or attendees actively recommending an artist to their friends.

Following the Last.fm ontology, we bifurcate music artists into two categories: Hyped and Top Artists.[1] Hyped Artists have the largest *increases* in listenership, while Top Artists have the highest play counts (Whiting, 2012). An example of a Top Artist in our dataset is Vampire Weekend, while the artist Yo-Yo Ma appears in our Hyped Artists list. This distinction is necessary for two reasons. From an economic perspective, industry insiders claim only top mainstream artists seem to turn a profit from touring (Passman, 2006). Second, a number of music sociologists have concluded that “mainstream” music is its own music scene[2] (e.g., Crossley and Emms, 2016). And while it would be interesting to see the concert-attendance effects in distinct musical niches (e.g., heavy metal, electronica, jazz), it is difficult to get sufficiently large sample sizes for each niche. After all, artists outside of the mainstream tend to have much lower concert attendance. As such, to get some estimate of concert-attendance effects of artists outside of the mainstream, we have to rely on an aggregate of niches. The Hyped Artist list allows us to include artists that are niche but have seen rapid increases in listenership indicating that they may soon become mainstream or they have, at the very least, gained popularity in their respective music scene. This ensures that we are including artists that are

[1] For detailed definitions of these and other terms used, please see the Methods section.

[2] Or “world” if you subscribe to the music worlds model.

sufficiently popular to have Last.fm users self-reporting attendance.

We extract live events between 2013 and 2014 documented on the Last.fm website; for each event, we extract the list of all the Last.fm users who reported attending. Then we extract the attendees' basic demographic information, as well as all the songs they listened to a month before and after that live event. We also capture their entire Last.fm friends list and the listenership records of each of those friends. This sampling strategy is a variant of "labelled star sampling" (Kolaczyk, 2009).

Data and methods

The analysis in this paper uses publicly available data from the music website Last.fm. We use three distinct types of data: (1) track listens, (2) event attendance, and (3) the Last.fm friends network.

Track listens data

Last.fm is a free music website with over 20 million active users (Weber, 2012) that keeps track of the songs played by its members. A user does not have to listen to the music directly from the Last.fm website for it to be recorded (or "scrobbled") in the user's track history—a user needs only to install the Audioscrobbler plugin on their music software (e.g., Itunes, Windows Media Player, etc.). The plugin keeps track of the music, as well as the time and date when the track is played—even when the user is not connected to the Internet. When the user next reconnects to the Internet, the stored data is dumped onto Last.fm's servers with date information updating retroactively. We extract this data to be used as our main outcome variable of interest.

In 2014, Last.fm partnered with Spotify (Dredge, 2014), which led to the inclusion of the Audioscrobbler plug-in in any Spotify installation by default, such that the user just needed to switch on Last.fm music scrobbling in their Spotify settings page.

The plugin and the API can be manipulated by users with programming experience to record tracks that someone has not actually listened to. We exclude cases where, in a given day, the individual appeared to have listened to a given artist for a duration that is greater than the length of a day. The average song length of contemporary popular music is 227 s (Anisko and Anderson, 2012) and so more than 380 songs in a day is unlikely to translate into actual listens.[3] Since we are unable to easily establish whether a Last.fm user is active, we apply our regression discontinuity only to users who listened to at least one song by the target artist in the 2-month observation window.

Event attendance data

Last.fm also serves as a platform for publicizing live events, which can be uploaded by fans, promoters, or the artists themselves. Last.fm listeners can indicate that they are going to attend or have attended an event. The event pages then retain the exact date and time of the event along with all the attendees. Events prior to the existence of Last.fm can also be uploaded (e.g. the band Deep Purple have events as far back as 1968 with 8 self-reported attendees). To avoid recall bias (i.e., where users will retroactively mark that they attended only the most memorable events), we concentrated on recent live events. In 2014, we extracted live events from "Top Artists" and "Hyped Artist" that occurred January, 2013–October, 2014. Hyped Artists correspond to bands who have the highest rate of growth while Top Artists are the most listened to artists on the site overall (Whiting, 2012).[4]

We aimed to get comparable sample sizes of attendees in the Top Artists and the Hyped Artists datasets, so we ended up extracting the live events of 85 Top Artists and 301 Hyped Artists. As expected, Top Artists had more documented live events and higher levels of self-reported attendance. We excluded artists that have not actually had live events in recent years (e.g. Queen). (We also excluded all Bob Dylan events as his page was associated with a number of misclassified live events; our results do not meaningfully change when we re-include Bob Dylan events.) We also ensured that no artists are found both in the Hyped and Top Artist datasets. We only included events that had one eligible Hyped or Top Artist playing.

"Friends" data

Last.fm also serves as a music-based social network; users are able to "friend" other people and discuss music and artists primarily through "shoutboxes." Past research indicated that users primarily friend other users with similar musical taste (Baym and Ledbetter, 2009), which makes this network particularly susceptible to indirect effects. But what exactly is the nature of Last.fm friendship ties? Bischoff (2012) found suggestive evidence indicating that a substantial number of Last.fm friendship links correspond to what are likely real-world social links. But even among users who are solely online friends, there is a substantial amount of public (reciprocal) communication (Bischoff, 2012). There is also some causal evidence that social influence does occur on Last.fm; Bapna and Umyarov (2015) conducted a randomized experiment in gifting users paid-premium subscription accounts, finding that this increased the odds their friends pay for premium by ~60%.

We extracted a network of friends for all attendees of events by Hyped and Top Artists. We suspect there were scraping errors for some attendees resulting in the extraction of an incomplete friends list. All analyses are robust to the exclusion of these suspect attendees. We then extracted two months of listening history (one month before and after the date of the live event) of the corresponding artist for each friend of each attendee.

Time-series regression discontinuity design

We use a regression discontinuity design as applied to time-series to establish causal identification (Malik and Pfeffer, 2016). In an RDD, a treatment is applied at a cutoff along some continuous running variable; the RDD estimator identifies the *instantaneous* impact at that cutoff, so long as the treatment (post-cutoff) and control (pre-cutoff) units are not able to select which side of the cutoff they are. The key advantage of this design in a time-series context is that live event treatment D_i does not necessarily have to be a random event for Last.fm user i. Rather, the *timing* of the event at time t_i has to be as-if random. In other words, a unit i can choose to go to event D_i and thus receive treatment, but that individual is unable to determine the time at which D occurs. While the individual clearly self-selects into attending the concert, because they (usually) cannot determine the time and date of the concert, applying RDD to this individual's timeline of $Y_{i,t}$ is acceptable so long as $(Y_{i,t}(0), Y_{i,t}(1))$ is independent of D_i. The one lingering question is whether D_i actually corresponds to concert attendance and no other variable. This is a reasonable concern which we address by investigating the potential impact of marketing effects around the event. The key feature of RDD designs is that they allow us to identify the instantaneous impact of a particular treatment D_i. While other variables may have impact on listenership $Y_{i,t}$, they would also have to occur precisely at the same point in time as the live event to bias our treatment effect estimates.

To be specific, our methodology for calculating direct effects is as follows. First, we take individual i and observe their listenership of artist a for three weeks before and after they attend a live event c by artist a. Applying the standard RDD framework (Imbens and Lemieux, 2008; Jacob et al., 2012),[5] the direct impact estimator is

[3] All analyses are robust to the inclusion of these outliers.

[4] We were not able to get official verification from Last.fm as to the length of the time period used in the calculation of the Hyped Chart, but from our investigations the timescale appeared to be a month.

$$\hat{\tau}_D = \frac{1}{N}(\lim_{b \to 0^+} E[Y_{i,t} \mid t_i = b] - \lim_{b \to 0^-} E[Y_{i,t} \mid t_i = b])$$

where b is the bandwidth generated using the leave-one-out cross validation approach to minimize squared bias and variance (Imbens and Kalyanaraman, 2012) and the cutoff at time zero is the live event of interest for individual i.

The estimation of indirect effects is slightly more complicated. Let A be the set of all i that have $D_{i,t} = 1$ for some value of t_k and $= 0$ for another value of t_j where $j \neq k$. These are all the nodes who were treated at their respective cutoff. Now let B be the set of all i for whom $D_{i,t} = 0$ for all t who have contiguous neighbours in A. The indirect effect estimator is thus

$$\hat{\tau}_{ID} = \frac{1}{M}(\lim_{b \to 0^+} E[Y_{i,t} \mid t_j = b] - \lim_{b \to 0^-} E[Y_{i,t} \mid t_j = b])$$

with t_i standardized across unit i's neighbour's cutoff c_j where j and i are neighbors and i is in set B, while $j \in A$. (M is the total number of nodes in set B.) However, this estimator is perhaps too broad, as it does not consider that we have different levels of dosage depending on where a node is located in the network. Specifically, an individual may have *many* friends who attended live event c and it is unlikely that the persuasive impact for that node will be the same as for an individual who has only *one* friend that went to live event c. Calculating the average effect for these varying levels of dosage is as trivial as calculating $\hat{\tau}_{ID}$ for only those individuals with a specific level of treatment exposure (i.e. a subgroup analysis). However, we cannot easily calculate standard errors using conventional methods. One means of addressing this challenge is by adopting the randomization inference approach as used by Cattaneo et al. (2015).

Following the standard for randomization inference approaches (Ho and Imai, 2006), we must randomly permute treatment assignment while holding constant $Y_{i,t}$ (and any $X_{i,t}$ covariates if available) *and* the structure of our social network. In the usual RD designs, we would have difficulties in holding the network structure constant while randomly permuting treatment assignment. In the case of RD as applied to time-series *within* units, we do not have the same issue. We can easily select an arbitrary value for the cutoff that does not correspond to any treatment event of that node's immediate neighbours. To be specific, the algorithm is as follows. For individual i in set B (i.e. any individual who did not attend a live event but had friends who did), we select another event by the same artist to be cutoff c_i. This event has to both occur in the two months of data scraped for individual i and not correspond to a cutoff c_j in set A. In other words, we select an event that occurred in the same period of time where we have data for individual i and none of individual i's Last.fm friends attended that event. We do this for all individuals with exactly one attendee-friend 40 times to produce 40 different synthetic null distributions. We then calculate the p-value by determining the number of times the observed effect occurred among our null distributions. We repeat this procedure for individuals with two attendee-friends and so on.

We report all results using the bandwidth generated using the leave-one-out cross validation approach to minimize squared bias and variance (Imbens and Kalyanaraman, 2012).[6] As recommended in Imbens and Lemieux (2008), we use the rectangular kernel for all analyses and verify its robustness using a triangular kernel. All analyses with the triangular kernel exhibit similar results. Since the same Last.fm user could have attended multiple events by a Top Artist in our dataset, we cluster standard errors on attendee. The results do not change meaningfully when we do not cluster the standard errors. Because each of these regression discontinuities used a slightly different bandwidth, as a robustness check, we use the Top Artist bandwidth with the Hyped Artist regression discontinuity and vice versa; the reported discontinuity impacts do not change meaningfully.

[5] We use the Stata package "rd" coded by Nichols (2011).

[6] Since the cutoff point is a theoretical concept, we must approximate the limits using some finite bandwidths. The closer to the cutoff that we are, the less bias we have, but we reduce our sample size by excluding more observations around the cutoff thereby decreasing efficiency. Imbens and Kalyanaraman (2012) simply automate the process of optimizing bias and variance.

Results

Identifying differences in the characteristics of Hyped and Top Artist event attendees helps build the case that music consumption and influence flows may be different across these two datasets. (Previous studies have found that fundamental demographic differences such as gender are associated with different motivations for attending live events (Bowen and Daniels, 2005).) As seen in Table 1, we find large differences in the gender composition of Hyped and Top Artist event attendees. Namely, Hyped Artists tend to attract ~54% of men to their shows, while Top Artists have a substantially more equitable gender breakdown with 49.7% of attendees reporting to be male. We should note that more than a tenth of our participants refrained from reporting their gender, so it is possible that it is not necessarily the actual gender composition of live events that differs across Top and Hyped bands, but rather the willingness to report a particular gender.

We then look at the friends-network of these attendees. As seen in Table 2, the local metrics of the two networks are relatively similar, with similar numbers of friends on average. Our sampling strategy precludes us from being able to cite any statistics about the global network structure—particularly metrics of transitivity such as betweenness (Kolaczyk, 2009). Fig. 1 illustrates one of the sub-networks that make up our data sample: a Metallica concert in Ecuador with all the self-reported Last.fm attendees (in blue) and their non-attending friends (in red).

Direct effects

We find strong evidence of direct impacts on listenership among concert attendees of both Top and Hyped Artists. As seen in Fig. 2, a Top Artist's live event increases listenership by 1.13 songs[7] (s.e. = 0.08, optimal bandwidth = 1.51, z-test p-value $<$ 0.001), while a Hyped Artist live event increases listenership by 1.05 of a song (s.e. = 0.08, optimal bandwidth = 1.87, z-test p-value $<$ 0.001).

Due to Last.fm's integration with Spotify, it is possible to estimate the impact of this increase in listenership on artist revenue.[8] Since Spotify states that an artist earns on average \$0.006 and \$0.0084 per stream (Spotify, n.d.), if we assume each play is the average of these two figures, or \$0.0072, we find that a Hyped Artist earns \$0.0076 per average attendee in additional revenue from a live event. Similarly, a Top Artist earns an additional \$0.0081 per average attendee. It is important to note that these impacts are likely much larger. As seen in Fig. 2, listenership increases before *and* after the event. For instance, within two days of Top Artist concerts, music listenership is on average 2.01 songs per day, while at all other points in our data, it is 0.55 songs per day; for Hyped Artists, it is 1.37 songs per day within two days of a concert and only 0.30 songs per day otherwise. These impacts immediately *before* the event may, for instance, be the result of promotional and marketing efforts. However, we must emphasize that pre-

[7] Since RDD identifies an *instantaneous* effect, we refrain from describing this impact in terms of songs *per day*, particularly because the bandwidth is often *more* than a day. As such, one could frame this effect as the increase in listenership *immediately* after the event. That said, the estimate is literally the difference between the song plays in the bandwidth time window *after* the event and the song plays in the bandwidth window *before* the event. If all assumptions hold, then this corresponds to the instantaneous impact of the event.

[8] Since the Last.fm song data includes not just Spotify song plays, but also users playing physical CDs and pirated material, the monetary estimates are used only to illustrate the magnitude of potential impacts.

Table 1
Demographic statistics of attendees.

Top artist dataset			Hyped artist dataset		
Artists in sample	85		Artists in sample	301	
Live events	2,237		Live events	4,344	
Attendees	27,137		Attendees	33,916	
Attendees self-reported gender			Attendees self-reported gender		
Female	9,958	36.70%	Female	10,641	31.37%
Male	13,475	49.66%	Male	18,323	54.02%
Not stated or missing	3,704	13.65%	Not stated or missing	4,586	13.52%
Attendees self-reported location			Attendees self-reported location		
United States	4,096	15.09%	United States	4,717	13.91%
United Kingdom	2,733	10.07%	United Kingdom	3,721	10.97%
Germany	1,749	6.45%	Germany	3,210	9.46%
Australia	1,465	5.40%	Poland	2,679	7.90%
Netherlands	1,021	3.76%	Netherlands	2,107	6.21%
Other, not stated, or missing	16,073	59.23%	Other, not stated, or missing	17,482	51.54%

Table 2
Friends network.

Top artist dataset		Hyped artist dataset	
Attendees & friends of attendees (node count)	624,194	Attendees & friends of attendees (node count)	732,192
Friend connections (edge count)	1,493,536	Friend connections (edge count)	1,714,049
Average number of friends of attendees (average degree)	56.8	Average number of friends of attendees (average degree)	52.6

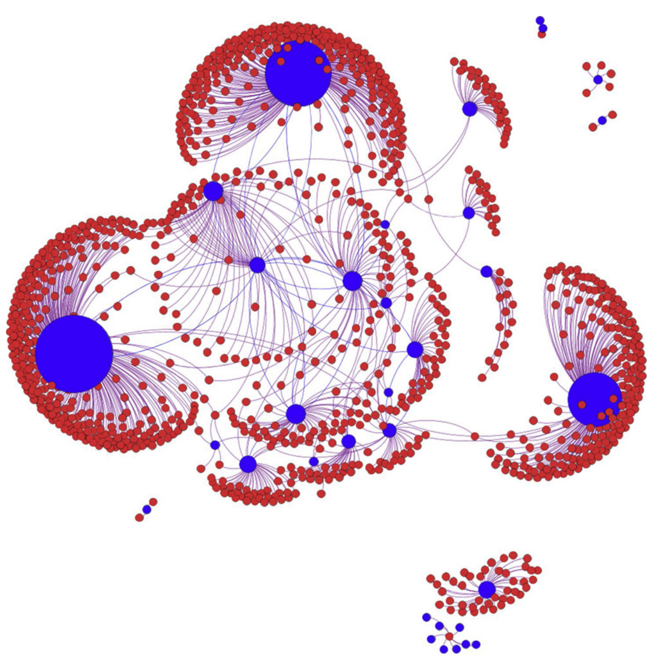

Fig. 1. Friends network visualization of attendees of Metallica concert in Quito, Ecuador on March 18, 2014. The blue nodes show the attendees and the red nodes show the non-attendees. (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

and post-event increases in listenership are not causally identified.

As an example, one event included in the dataset of Top Artist events is a Taylor Swift concert at the O2 Arena in London, which had a reported attendance of 74,740 (The Red Tour, n.d.). This means that if 66%[9] of those attendees were Spotify listeners, this event generated an additional $401.34 from attendees' subsequent song listens.

Indirect effects

We also evaluate if the attendees influence their friends, who have *not* attended the same event. Therefore, we apply the same regression discontinuity design to all friends of the attendees. One important note: to ensure we are looking at active users, we include only those users who have listened to the artist at least once in the 2-month observation window. As illustrated in Fig. 3, we find a statistically significant impact on the listenership of the non-attending friends of attendees of Top Artists (0.060 songs, s.e. = 0.012, optimal bandwidth = 1.27, $p < 0.001$), but not Hyped Artists (.016 songs, s.e. = 0.025, optimal bandwidth = 1.91, $p > 0.05$). The indirect effect on friends of Top Artist attendees is a 0.060 additional song plays, or more than 5% of the direct impact on listenership.[10] While this is a trivial impact on its own, it is important to emphasize that the mean Top Artist attendee has 56.8 friends.[11] This means that one user's attendance translates to 3.4 more song plays on average, which translates to an increase of $0.025 per attendee. Using our earlier Taylor Swift concert example, the London concert secured an additional $1,210.40 in song streams. However, this analysis assumes only one friend attended the live event. We proceed by investigating whether the indirect impact increases as the number of friends who attended the event increases.

Since the data includes non-attendees who are friends with multiple attendees, we can pursue a series of subgroup analyses. A key limitation of our analysis is that event attendance is self-reported. But this potential measurement error could go in both directions.

To determine whether the number of attendee-friends interacts with the indirect effect, we run a series of regression discontinuity analyses across non-attending users with various numbers of attendee-friends. We find that as the number of friends who attended the event increases, the influence on listenership increases monotonically. To better discern if this pattern stems from multiple attendees exerting influence on the non-attender, we perform a permutation test. Namely, we look at the discontinuity estimates across the number of attendee-friends in our actual dataset and then compare these results to a synthetic dataset where we held the friend network constant but randomly assigned that non-attending friend to a different live event date of the same artist in the same 30-day period such that no user in their friend network attended that live event.[12] For more details on this procedure, please see the Methods section. Table 3 and Fig. 4 illustrate these results.

Even with the permutation test, we are not able to ascertain if there

[9] We base this estimate on recent proprietary studies. A recent Spotify study found more than 2/3rds of attendees of a Dutch festival in 2014 used Spotify (Page, 2014), while Music Watch Inc. found that 56% of Internet users streamed music in 2012 and 69% in 2014 (Crupnick, 2015).

[10] Non-attendees listen to the artist at an average of 0.47 songs per day within 2 days of the event and 0.41 at all other points in the dataset.

[11] If some friends-lists are incomplete, this would bias our estimates of indirect economic effects downward.

[12] We use the same bandwidth generated for the main indirect effects analysis to make comparisons across subgroups. Our analysis is robust to using a leave-one-out cross validation approach separately for each subgroup, though the results are weaker.

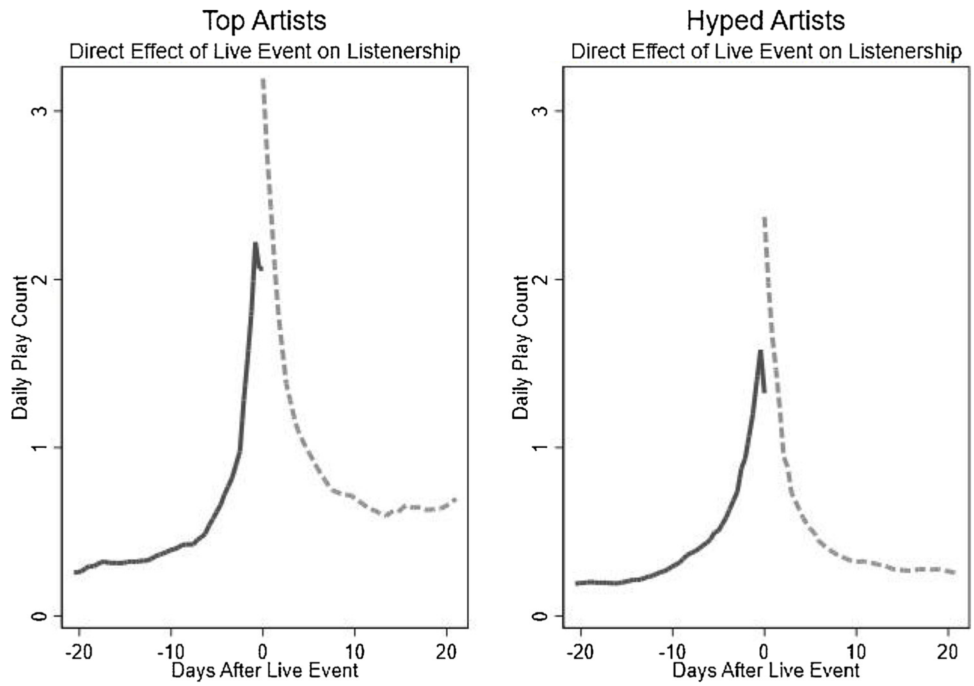

Fig. 2. Graphical depiction of the regression discontinuity estimate of the direct impact of live event.

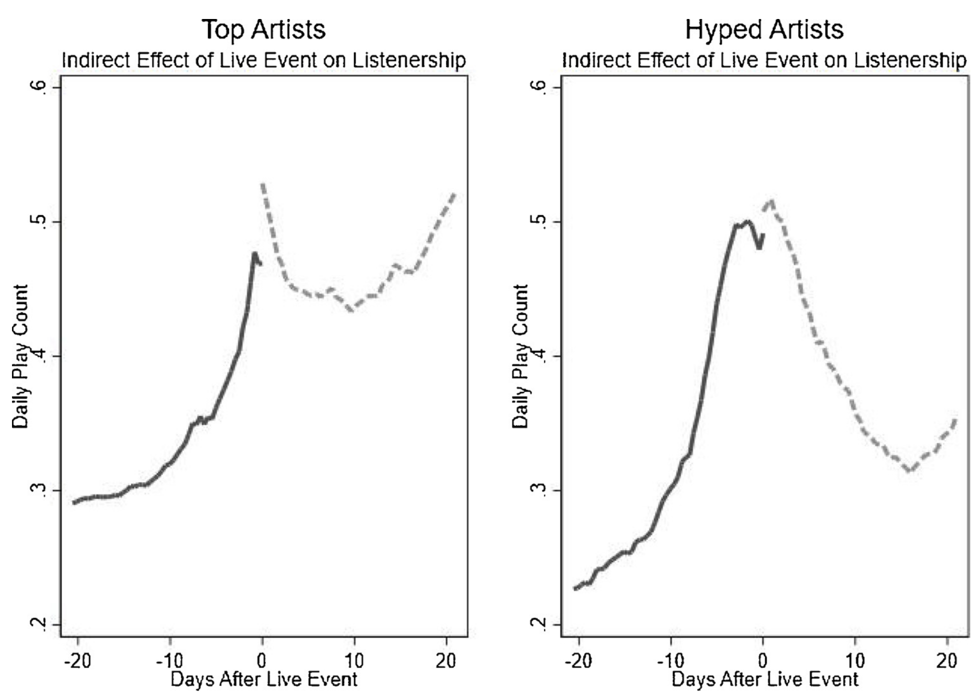

Fig. 3. Graphical depiction of the regression discontinuity estimate of the indirect impact of live event.

is a *causal* relationship between the number of friends and the increasing indirect effect. It is quite possible that having multiple friends attend a given live event is indicative of the level of marketing/advertisements about the event (i.e. more marketing may induce more members of a given social circle to attend an event). However, the impact of marketing is unlikely to be tied to the day of the event. Rather, marketing about a concert is much more likely to occur in the days leading up to the event, which may explain the rapid increases in listenership immediately before our discontinuities.

Subgroup analysis

Since prior studies have shown that the motivations for live event attendance (Bowen and Daniels, 2005; Pegg and Patterson, 2010) and music listenership patterns (Park et al., 2019*)* tend to differ across standard demographic characteristics such as gender, we check whether the impact of live events on music consumption differs across location and gender.[13] These hypotheses were not made *a priori,* and so due to

[13] While users can also share their age, this data was not available on Last.fm's API at the time this analysis was conducted.

Table 3
Permutation test results.

# Friends who attended event	Observed effect	Null distribution average	Permutation test p-value (one-sided)
1	0.033	0.005	< 0.025
2	0.164	−0.008	< 0.025
3	0.493	0.041	< 0.025
4	1.258	0.276	< 0.025
5	1.639	−0.243	< 0.025

We exclude all instances of non-attendees with more than 5 friends due to the likelihood that the non-attending friend may have simply failed to report attendance. (See for instance the cluster of attendees in the lower right corner in Fig. 2.) Less than 0.2% of our sample have more than 5 friends who attended the same event.

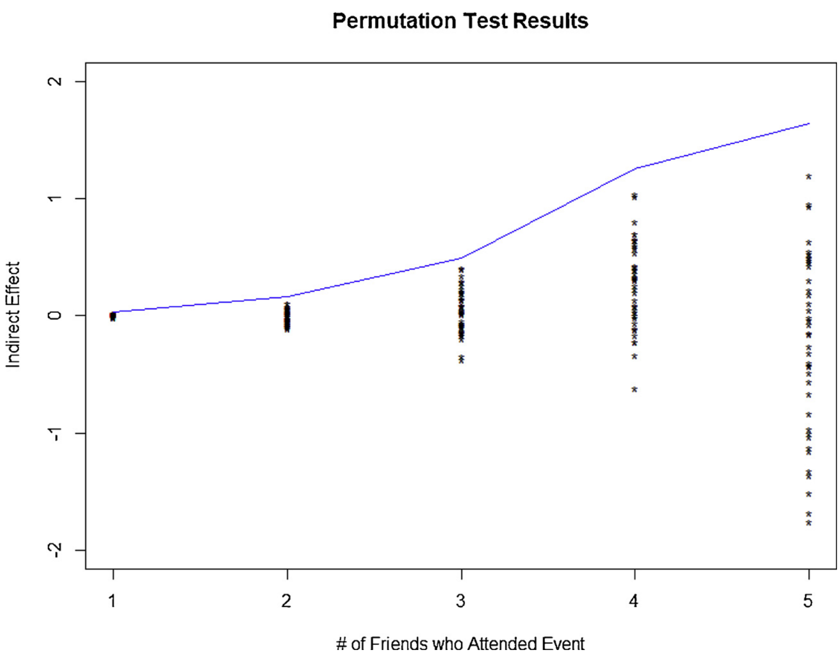

Fig. 4. Graphical depiction of permutation test results.
The blue line indicates the observed indirect effect among each subgroup on the x-axis. The asterixis correspond to a randomly permuted null distribution, where the network remains the same, but the cutoff is randomly permuted to another event by the same artist that was not attended by any of that individual's friends. (For interpretation of the references to colour in the figure legend, the reader is referred to the web version of this article.)

Table 4
Impacts on listenership across attendee demographics measured by daily song plays.

Direct Effects and Indirect Effect							
Hyped artists				Top artists			
Male	Female	US	non-US	Male	Female	US	non-US
0.47	0.58	0.91	0.96	1.26	0.92	1.11	1.07
0.036	0.054	−0.002	0.028	0.054	0.051	0.027	0.066

the risk of data-dredging (Assmann et al., 2000) we did not run any statistical tests across groups. Rather, we hope to get a Bayesian baseline for potential differences, which should be evaluated rigorously in future studies. Because there are clear differences in demographic characteristics between the attendees of Top and Hyped Artist shows, we pursue subgroup analysis separately for each batch of data (Table 4). While there are some slight differences in the direct impact of Top Artist live events on males as compared to females, males have greater direct impacts of concert attendance. However, we should not make any rigorous comparisons of these statistics, as there may be large biases in gender self-reporting.

We also investigated whether the demographics of non-attending users translate to differential indirect effects (Table 4). In this case, we find that there are larger indirect impacts on non-US, non-attending users for both Hyped and Top Artists. There is a slightly larger indirect effect on female non-attending friends in the Hyped Artist universe, but it is important to note that the indirect effects in the Hyped Artist universe were not significant.

Discussion and conclusion

In this paper we leverage a time-series RD design to causally identify the impact of live events on music listenership. Because RDD identifies the instantaneous effect, the main threat to causal inference is the possibility that some other variables correspond perfectly with the event itself. The only other variable that could conceivably perfectly correspond to the live event and affect music listenership are marketing campaigns promoting that live event. However, this would mean that marketing efforts would need to reach their zenith on the day of the event. Qualitative reports from the music business seem to imply that this is unlikely (Rogers, 2013). We believe our analysis provides sufficient evidence of peer influence. Namely, users who had friends on the Last.fm network were exposed to both an attending-user's increased listenership of a given band and their attendance. However, in all cases, the exhibition of that information was not as obtrusive and explicit as, for instance, Facebook's newsfeed. This should reduce the magnitude of peer effects. Additionally, users may have communicated either through public shoutboxes or personal messages about the events. As seen in Bischoff (2012) many users may be familiar with their Last.fm friends in real life and on other social media platforms; as such, they may give music recommendations outside of the Last.fm network. We believe that the indirect effects were some combination of these means of communication.

The results illustrate that even without any profits from tickets and merchandise sales at shows, there may be long-term economic benefits for touring bands. There are sizable, statistically significant impacts on music consumption of attending a live event. The size of the *direct* impact is comparable for both Top Artists and Hyped Artists. However, while there are substantial indirect effects for Top Artists, there are no statistically significant effects for the up-and-coming artists. In concordance with the Black et al. (2007) study, we see that the rich get richer, while the average Hyped Artist struggles to expand their fanbase through touring. The necessary conditions for contagion to occur are twofold: (1) the artist must be popular and (2) at least four Last.fm friends should have attended that artist's live event for "infection" to occur. (By infection, we mean peer influence that has at least as large an effect on listenership behaviour as attending a live event). Our most noteworthy finding is that if enough of your friends see a live event, there is a bigger boost in listenership of that artist than even if you yourself attended the event. This suggests that the word-of-mouth effects have the potential to be more important in indirect revenues than perhaps even direct ticket sales. That said, it is important to emphasize that the number of co-attending friends is not randomly assigned and, so there may be some unobserved variable that is responsible for the difference in contagion effects.

One important consideration is physical geography since the geographical availability of live shows imposes additional restrictions on concert attendance. Aside from the actual cost of a concert ticket, a fan may need to incur substantial costs to travel to the closest venue where their favourite artist is playing. This may have two implications. First, individuals who travel to see an artist may have a greater incentive to recommend the artist to others. To offset the financial cost, they may be much more enthusiastic in talking about the event with friends since gig "collection" is often a part of musicking (Crossley and Bottero, 2015b). Second, individuals who want to see a particular artist but are unable to travel may be more prone to have their listenership influenced by online friends who did see the artist in their town. Indeed, individuals who live in more remote locations may be much more dependent on online recommendations to discover new artists.

In the developments of the economic geography of music, we have seen an increasing concentration of music scenes in major metropolitan hubs (Florida et al., 2010), so we are much more likely to have clusters of Last.fm users co-attending the same concert in major cities such as Los Angeles and New York City and fewer users co-attending in more remote locations. As such, one ambiguity of our contagion effect is whether users in metropolitan areas are simply more influential than users in smaller towns. Even if this is the case, the contagion effect is nevertheless an important marker for artists and justifies playing shows in places where there are high levels of co-attendance. High levels of co-attendance may also correspond to a tighter-knit music scene.

Future studies should also analyse just how much listenership increases across all (monetized) streaming platforms. Our results only cover those individuals who use Last.fm and have enabled music tracking on their preferred streaming service. These individuals may be systematically different from the usual concert attendee/streaming-music-listener.

Acknowledgments

TY was partially supported by The Alan Turing Institute under the EPSRC grant EP/N510129/1. The Funder had no role in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the manuscript.

References

Anisko, N., Anderson, E., 2012. Average Length of Top 100 Songs on iTunes. Retrieved from:. StatCrunch. http://www.statcrunch.com/5.0/viewreport.php?reportid=28647&groupid=948.
Aral, S., Nicolaides, C., 2017. Exercise contagion in a global social network. Nat. Commun. 8, 14753. https://doi.org/10.1038/ncomms14753.
Assmann, S.F., Pocock, S.J., Enos, L.E., Kasten, L.E., 2000. Subgroup analysis and other (mis) uses of baseline data in clinical trials. Lancet 355 (9209), 1064–1069. Retrieved from: http://www.sciencedirect.com/science/article/pii/S0140673600020390http://www.sciencedirect.com/science/article/pii/S0140673600020390.
Bapna, R., Umyarov, A., 2015. Do your online friends make you pay? A randomized field experiment on peer influence in online social networks. Manage. Sci. 1902–1920.
Basse, G.W., Airoldi, E.M., 2018. Model-assisted design of experiments in the presence of network-correlated outcomes. Biometrika 105 (4), 849–858.
Baym, N.K., Ledbetter, A., 2009. Tunes that bind? Predicting friendship strength in a music-based social network. Inf. Commun. Soc. 12 (3), 408–427.
Bischoff, K., 2012. We love rock'n'roll: analyzing and predicting friendship links in Last.fm. In: Proceedings of the 4th Annual ACM Web Science Conference. June. ACM. pp. 47–56.
Black, G.C., Fox, M.A., Kochanowski, P., 2007. Concert tour success in North America: an examination of the top 100 tours from 1997 to 2005. Pop. Music Soc. 30 (2), 149–172.
Bond, R.M., Fariss, C.J., Jones, J.J., Kramer, A.D., Marlow, C., Settle, J.E., Fowler, J.H., 2012. A 61-million-person experiment in social influence and political mobilization. Nature 489 (7415), 295–298.
Bowen, H.E., Daniels, M.J., 2005. Does the music matter? Motivations for attending a music festival. Event Manage. 9 (3), 155–164.
Cattaneo, M.D., Frandsen, B.R., Titiunik, R., 2015. Randomization inference in the regression discontinuity design: an application to party advantages in the US Senate. J. Causal Inference 3 (1), 1–24.
Charron, J.P., 2017. Music audiences 3.0: concert-goers' psychological motivations at the dawn of virtual reality. Front. Psychol. 8.
Centola, D., 2010. The spread of behavior in an online social network experiment. Science 329 (5996), 1194–1197.
Cohen-Cole, E., Fletcher, J.M., 2008a. Detecting implausible social network effects in acne, height, and headaches: longitudinal analysis. BMJ 337, a2533.
Cohen-Cole, E., Fletcher, J.M., 2008b. Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic. J. Health Econ. 27 (5), 1382–1387.
Crossley, N., Bottero, W., 2015a. Social spaces of music: introduction. Cult. Sociol. 9 (1), 3–19.
Crossley, N., Bottero, W., 2015b. Music worlds and internal goods: the role of convention. Cult. Sociol. 9 (1), 38–55.
Crossley, N., Emms, R., 2016. Mapping the musical universe: a blockmodel of UK music festivals, 2011–2013. Methodol. Innov. 9, 2059799116630663.
Crossley, N., McAndrew, S., Widdop, P. (Eds.), 2014. Social Networks and Music Worlds. Routledge.
Crupnick, R., 2015. From Stream to Ticket: Mapping the Value of Music Discovery. Retrieved from:. Music Watch Inc.. http://www.musicwatchinc.com/blog/from-stream-to-ticket-mapping-the-value-of-music-discovery/.
Dewenter, R., Haucap, J., Wenzel, T., 2012. On file sharing with indirect network effects between concert ticket sales and music recordings. J. Media Econ. 25 (3), 168–178.
Dredge, S., 2014. Last.fm plots streaming music comeback with Spotify and YouTube. The Guardian January 30. Retrieved from: http://www.theguardian.com/technology/2014/jan/30/lastfm-streaming-comeback-spotify-youtube.
Earl, P.E., 2001. Simon's travel theorem and the demand for live music. J. Econ. Psychol. 22 (3), 335–358.
Florida, R., Mellander, C., Stolarick, K., 2010. Music scenes to music clusters: the economic geography of music in the US, 1970–2000. Environ. Plan. A 42 (4), 785–804.
Hahn, J., Todd, P., Van der Klaauw, W., 2001. Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica 69 (1), 201–209.
Ho, D.E., Imai, K., 2006. Randomization inference with natural experiments: an analysis of ballot effects in the 2003 California recall election. J. Am. Stat. Assoc. 101 (475), 888–900.
Imbens, G., Kalyanaraman, K., 2012. Optimal bandwidth choice for the regression discontinuity estimator. Rev. Econ. Stud. 79 (3), 933–959.
Imbens, G.W., Lemieux, T., 2008. Regression discontinuity designs: a guide to practice. J. Econom. 142 (2), 615–635.
Jacob, R.T., Zhu, P., Somers, M.A., Bloom, H.S., 2012. A Practical Guide to Regression Discontinuity. MDRC.
Jones, J.J., Bond, R.M., Bakshy, E., Eckles, D., Fowler, J.H., 2017. Social influence and political mobilization: further evidence from a randomized experiment in the 2012 U.S. presidential election. PLoS One 12 (4), e0173851. https://doi.org/10.1371/journal.pone.0173851.
Kolaczyk, E., 2009. Statistical Analysis of Network Data. Springer.
Krueger, A.B., 2005. The economics of real superstars: the market for rock concerts in the material world. J. Labor Econ. 23 (1), 1–30.
Leskovec, J., Adamic, L.A., Huberman, B.A., 2007. The dynamics of viral marketing. ACM Trans. Web 1 (1), 5.
Liebowitz, Stan, 2003. Will MP3 downloads annihilate the record industry? The evidence so far. In: Libecap, Gary (Ed.), Advances in the Study of Entrepreneurship, Innovation, and Economic Growth. JAI Press.
Maasø, A., 2016. Music streaming, festivals, and the eventization of music. Pop. Music Soc. 1–22.
Rogowski, J.C., Sinclair, B., 2012. Estimating the causal effects of social interaction with endogenous networks. Political Anal. 20 (3), 316–328.
Malik, M.M., Pfeffer, J., 2016. Identifying platform effects in social media data. Tenth International AAAI Conference on Web and Social Media March.
Mark, N., 1998. Birds of a feather sing together. Soc. Forces 77 (2), 453–485.
Mark, N.P., 2003. Culture and competition: homophily and distancing explanations for cultural niches. Am. Sociol. Rev. 319–345.
Mønsted, B., Sapieżyński, P., Ferrara, E., Lehmann, S., 2017. Evidence of complex contagion of information in social media: an experiment using Twitter bots. PLoS One 12 (9), e0184148. https://doi.org/10.1371/journal.pone.0184148.
Montoro-Pons, J.D., Cuadrado-García, M., 2011. Live and prerecorded popular music consumption. J. Cult. Econ. 35 (1), 19–48.
Nichols, Austin, 2011. rd 2.0: Revised Stata Module for Regression Discontinuity Estimation. Retrieved from:. http://ideas.repec.org/c/boc/bocode/s456888.html.
Page, W., 2014. Adventures in the Lowlands: Spotify, Social Media and Music Festivals. Retrieved from:. Spotify Insights. https://insights.spotify.com/us/2014/10/22/adventures-in-the-lowlands/.
Park, M., Thom, J., Mennicken, S., Cramer, H., Macy, M., 2019. Global music streaming data reveal diurnal and seasonal patterns of affective preference. Nat. Hum. Behav. 3 (3), 230.
Passman, D.S., 2006. All You Need to Know About the Music Business, 6th ed. Free Press.

Pegg, S., Patterson, I., 2010. Rethinking music festivals as a staged event: gaining insights from understanding visitor motivations and the experiences they seek. J. Conv. Event Tour. 11 (2), 85–99 Taylor & Francis Group.

Resnikoff, Paul, 2017. Live concerts + streaming = 73% of the US music industry. Digital Music News June 7. Retrieved from: https://www.digitalmusicnews.com/2017/06/07/music-industry-concerts-streaming/.

Rodriguez, M.A., Gintautas, V., Pepe, A., 2008. A Grateful Dead analysis: the relationship between concert and listening behavior. First Monday 14 (1).

Rogers, J., 2013. The Death and Life of the Music Industry in the Digital Age. Bloomsbury Academic, London.

Romero, D.M., Meeder, B., Kleinberg, J., 2011. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. Proceedings of the 20th International Conference on World Wide Web 695–704 March. ACM.

Salganik, M.J., Dodds, P.S., Watts, D.J., 2006. Experimental study of inequality and unpredictability in an artificial cultural market. Science 311 (5762), 854–856.

Salganik, M.J., Watts, D.J., 2008. Leading the herd astray: an experimental study of self-fulfilling prophecies in an artificial cultural market. Soc. Psychol. Q. 71 (4), 338–355.

Small, C., 1998. Musicking: The Meanings of Performing and Listening. Wesleyan University Press.

Spotify. (n.d.). How is Spotify contributing to the music business? Retrieved from: http://www.spotifyartists.com/spotify-explained/#how-we-pay-royalties-overview.

The Red Tour, (n.d). In Wikipedia. Retrieved 27 Spetember 2015, from https://en.wikipedia.org/wiki/The_Red_Tour.

Washenko, A., 2017. PwC entertainment and media outlook, part 2: United States results. RAIN News June 8. Retrieved from: http://rainnews.com/pwc-entertainment-and-media-outlook-part-2-united-states-results/.

Weber, H., 2012. Spotify announces 5M+ paid subscribers globally, 1M paid in US, 20M total active users, 1B playlists. TNW December 6. Retrieved from: https://thenextweb.com/insider/2012/12/06/spotify-announces/.

Whiting, D., 2012. So This Is Christmas, and What Have We Done? Last.fm.