

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Exatas
Programa de Pós-Graduação em Ciência da Computação

Guilherme Soares Simões dos Santos

**Explorando a dinâmica de popularidade de músicas sampleadas, regravadas
e remixadas**

Belo Horizonte
2025

Guilherme Soares Simões dos Santos

**Explorando a dinâmica de popularidade de músicas sampleadas, regravadas
e remixadas**

Versão Final

Dissertação apresentada ao Programa de Pós-Graduação em
Ciência da Computação da Universidade Federal de Minas
Gerais, como requisito parcial à obtenção do título de Mestre
em Ciência da Computação.

Orientador: Flávio Vinícius Diniz de Figueiredo

Belo Horizonte
2025

2025, Guilherme Soares Simões dos Santos.
Todos os direitos reservados

Santos, Guilherme Soares Simões dos.

S237e Explorando a dinâmica de popularidade de músicas sampleadas, regravadas e remixadas [recurso eletrônico] / Guilherme Soares Simões dos Santos. Belo Horizonte – 2025. 1 recurso online (109 f. il., color.) : pdf.

Orientador: Flávio Vinícius Diniz de Figueiredo.

Dissertação (Mestrado) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Ciência da Computação.

Referências: f. 78-80.

1. Computação – Teses. 2. Recuperação da informação – Teses. 3. Banco de dados – Música – Teses. 4. Direitos autorais – Música – Teses. 5. Música – Influência – Teses. I. Figueiredo, Flávio Vinícius Diniz de. II. Universidade Federal de Minas Gerais Instituto de Ciências Exatas, Departamento de Ciência da Computação. III. Título.

CDU 519.6*95(043)

Ficha catalográfica elaborada pela bibliotecária Irénquer Vismeg Lucas Cruz
CRB 6/819 - Universidade Federal de Minas Gerais - ICEX



UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

FOLHA DE APROVAÇÃO

Explorando a dinâmica de popularidade de músicas sampleadas, regravadas e remixadas

GUILHERME SOARES SIMÕES DOS SANTOS

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

PROF. FLAVIO VINICIUS DINIZ DE FIGUEIREDO - Orientador
Departamento de Ciência da Computação - UFMG

PROFA. ANA PAULA COUTO DA SILVA
Departamento de Ciência da Computação - UFMG

PROF. CARLOS HENRIQUE GOMES FERREIRA
Departamento de Computação e Sistemas - UFOP

PROFA. MIRELLA MOURA MORO
Departamento de Ciência da Computação - UFMG

Belo Horizonte, 31 de janeiro de 2025.

Ao meu avô “Cachimbeiro”, meu primeiro mestre e amigo.

Agradecimentos

Agradeço profundamente à minha família, cujo apoio foi essencial ao longo desta jornada. Em especial, guardo com carinho as palavras simples, porém sábias, de meu avô. Quando, ainda criança, reclamei que era impossível tocar um acorde CM7 no teclado por causa das minhas pequenas mãos, ele respondeu com uma frase que levo comigo até hoje: “O difícil fazemos agora, o impossível demoramos mais um pouco.” Essa lição de perseverança tornou-se um guia constante diante dos desafios.

Minha gratidão também à Fundação Mendes Pimentel (FUMP), que me permitiu continuar estudando ao oferecer o apoio necessário para minha alimentação e permanência na universidade durante a graduação; ao professor Flávio, por abrir as portas para minhas primeiras experiências científicas; e à Universidade Federal de Minas Gerais, por todas as vivências e aprendizados que transformaram minha vida, proporcionando os conhecimentos necessários para realizar este trabalho e construir um futuro melhor.

*“Thyself so reverence as to prefer the native growth of thy own mind to the richest
import from abroad; such borrowed riches make us poor.”*
(Edward Young, *Conjectures on Original Composition*)

Resumo

A digitalização transformou a produção, distribuição e consumo de música, introduzindo novas formas de composição, entre as quais se destacam os ‘empréstimos musicais’ (musical borrowings), onde compositores e produtores utilizam elementos de músicas pré-existentes para criar novas obras. Essas práticas geram controvérsias jurídicas, especialmente no que tange aos direitos autorais e ao uso justo, com debates sobre o impacto desses empréstimos no mercado da obra original. Alguns estudos sugerem que, em certos casos, os empréstimos musicais podem impulsionar as vendas das obras originais, levantando discussões sobre a possibilidade de considerar tais práticas como uso justo.

Diante desse cenário, a hipótese central deste trabalho é que os empréstimos musicais aumentam a popularidade das músicas originais. Para investigar essa questão, analisamos como diferentes formas de empréstimos musicais — como samples, covers e remixes — impactam a popularidade das músicas envolvidas. Utilizando dados do WhoSampled, que mapeiam aproximadamente 700 mil empréstimos, aplicamos um filtro para evitar falsos positivos, garantindo a qualidade do conjunto analisado. Em seguida, exploramos duas abordagens complementares: (i) uma análise estática, que compara os números absolutos de visualizações no YouTube e as métricas de popularidade no Spotify; e (ii) uma análise dinâmica, que investiga como os empréstimos musicais influenciam o interesse de busca pelas músicas originais ao longo do tempo. Para mensurar esses impactos, aplicamos o Regression Discontinuity Design (RDD) para avaliar os efeitos no curto prazo e a Causalidade de Granger para examinar os impactos de longo prazo.

Os resultados indicam que os empréstimos musicais impactam significativamente a popularidade das músicas originais. A análise com Regression Discontinuity Design (RDD) sugere que o lançamento de novas versões pode reavivar o interesse por músicas mais antigas. A Causalidade de Granger revela que, em 64% dos casos, as versões derivadas influenciam significativamente as músicas originais ao longo do tempo. A análise estática mostra que, embora as músicas originais sejam, em geral, mais populares, as versões derivadas apresentam pequenas diferenças de popularidade no Spotify, sugerindo a exploração da popularidade da obra original. Esses resultados oferecem novas perspectivas sobre a dinâmica econômica dos empréstimos musicais e podem contribuir para uma compensação mais justa no mercado musical.

Palavras-chave: web scraping; música; sampling; popularidade; influência.

Abstract

The digital era has transformed music production, distribution, and consumption, introducing new forms of composition, among which “musical borrowings” stand out. In these practices, composers and producers reuse elements from pre-existing songs to create new works. However, such practices spark legal controversies, particularly regarding copyright and fair use, raising debates about their impact on the market for the original works. Some studies suggest that, in certain cases, musical borrowings can boost the sales of the original works, fueling discussions about whether these practices could qualify as fair use.

In this context, the central hypothesis of this study is that musical borrowings increase the popularity of original songs. To investigate this, we analyze how different types of borrowings — such as samples, covers, and remixes — impact the popularity of the songs involved. Using data from WhoSampled, which maps approximately 700,000 borrowings, we applied a filter to avoid false positives, ensuring the quality of the dataset. Next, we explored two complementary approaches: (i) a static analysis comparing absolute YouTube view counts and Spotify popularity metrics, and (ii) a dynamic analysis investigating how borrowings influence search interest in the original songs over time. To measure these impacts, we employed Regression Discontinuity Design (RDD) to evaluate short-term effects and Granger Causality to examine long-term impacts.

The results indicate that musical borrowings significantly impact the popularity of original songs. The Regression Discontinuity Design (RDD) analysis suggests that the release of new versions can revive interest in older songs. Granger Causality reveals that, in 64% of cases, derivative versions significantly influence the original songs over time. Static analysis shows that while original songs are generally more popular, derivative versions exhibit small differences in Spotify popularity, suggesting an exploration of the original work’s popularity. These findings offer new perspectives on the economic dynamics of musical borrowings and could contribute to fairer compensation within the music market.

Keywords: web scraping; music; sampling; popularity; influence.

Lista de Figuras

1.1	Site WhoSampled.com exibindo informações sobre conexões musicais da obra “Crazy in Love” de Beyoncé. Fonte: WhoSampled.com	21
3.1	Evolução temporal do interesse de busca por músicas	29
3.2	Exemplo da aplicação da correlação de Pearson em séries temporais, com evidências de forte associação linear positiva.	30
3.3	Comparação entre DTW e distância euclidiana. Para fins de visualização, as séries temporais estão deslocadas verticalmente. Fonte: Tavenard [2021] . . .	31
3.4	Ilustração de causalidade de Granger entre as séries temporais X e Y . Padrões de X são repetidos em Y após um atraso temporal, conforme indicado pelas setas. Fonte: BiObserver [2025].	34
3.5	Representação gráfica do Regression Discontinuity Design. A figura mostra os grupos de controle e tratamento, destacando o efeito médio do tratamento (ATE) no ponto de corte. Fonte: Liu et al. [2022].	35
4.1	Fluxo do processo de coleta e integração dos dados. A figura destaca as fontes utilizadas e os métodos empregados para responder às perguntas de pesquisa.	37
4.2	Exemplo de um empréstimo musical do tipo “sampled” entre as músicas S e T.	38
4.3	Captura de Tela do Google Trends exibindo a ferramenta de análise de interesse ao longo do tempo.	40
4.4	Exemplo de resultado da busca do Google exibindo informações da base de conhecimento para “Michael Jackson”, incluindo sua data de nascimento, links para plataformas de música e conexões relevantes.	41
5.1	Distribuição da similaridade de strings normalizada para o nome da música .	47
5.2	Distribuição das músicas pelo ano de publicação no YouTube.	48
5.3	Distribuição do número de visualizações das músicas. A escala do eixo y é logarítmica e representa a porcentagem de músicas.	48
5.4	Distribuição dos empréstimos musicais em que a música derivada supera ou não a original em visualizações no YouTube, agrupados por ano de publicação.	49
5.5	Função de Distribuição Cumulativa Empírica (ECDF) para as diferenças percentuais simétricas.	50
5.6	Distribuição de visualizações de músicas originais e derivadas por tipo de empréstimo musical.	51
5.7	Distribuição das músicas por ano de lançamento.	52

5.8	Popularidade média das músicas por ano de lançamento.	52
5.9	Distribuição das músicas pela métrica de popularidade do Spotify.	53
5.10	Comparação entre a popularidade da música original e suas versões ao longo do tempo, no Spotify.	53
5.11	Função de Distribuição Acumulada da diferença percentual absoluta de popularidade entre a música original e a versão.	54
5.12	Proporção de músicas originais e versões mais populares por tipo de empréstimo musical.	54
6.1	Sugestões exibidas ao digitar o termo “Clube da Esquina”. A seleção da primeira opção realiza uma busca por termo, enquanto todas as outras realizam uma busca por entidade. Fonte: Google Trends — https://trends.google.com/trends	58
6.2	Distribuição da similaridade de strings normalizada entre informações do Wikidata e o conjunto original	62
6.3	Script de coleta das séries temporais utilizando Pytrends	63
6.4	Distribuição do coeficiente de correlação de Pearson entre as séries temporais de interesse para músicas originais e suas versões.	65
6.5	Distribuições das medidas de similaridade entre séries temporais de músicas originais e suas versões.	66
6.6	Distribuição do Efeito Médio do Tratamento (ATE) log-transformado para os 82 empréstimos musicais com significância estatística.	68
6.7	Interesse de busca por “ <i>Gotta Go Home</i> ” de Boney M.	69
6.8	Interesse de busca por <i>Calabria</i> de Rune RK	70
6.9	Exemplos de ATE	71
6.10	Interesse de busca por “ <i>Stronger</i> ” de Britney Spears	71
6.11	Distribuição dos p -valores do teste de Granger	72
6.12	Interesse de busca para músicas com alta relação causal	73
6.13	Interesse de busca para músicas recentes com alta relação causal	73
B.1	Kryptonite (3 Doors Down, 1999) sampleado em Superhuman (Chris Webby, 2014). ATE Normalizado: -3.11 — Causalidade de Granger: 0.32	82
B.2	It Takes Two (Rob Base & DJ E-Z Rock, 1988) sampleado em Peaches N Cream (Snoop Dogg song) (Snoop Dogg, 2015). ATE Normalizado: 4.451 — Causalidade de Granger: 0.18	82
B.3	U Can’t Touch This (MC Hammer, 1990) sampleado em Dance (Ass) (Big Sean, 2011). ATE Normalizado: 6.025 — Causalidade de Granger: 0.84	83
B.4	Can You Feel It (The Jacksons, 1980) sampleado em Sorry (Madonna, 2005). ATE Normalizado: 5.271 — Causalidade de Granger: 0.75	83
B.5	Get Up Offa That Thing (James Brown, 1976) sampleado em Luv (Apink, 2014). ATE Normalizado: 3.809 — Causalidade de Granger: 0.84	83

B.6	Bon, Bon (Pitbull, 2010) sampleado em We No Speak Americano (Yolanda Be Cool, 2010). ATE Normalizado: -6.056 — Causalidade de Granger: 0.27	84
B.7	Real Love (Mary J. Blige, 1992) sampleado em Super Rich Kids (Frank Ocean, 2012). ATE Normalizado: 4.366 — Causalidade de Granger: 0.99	84
B.8	Get On Your Boots (U2, 2009) sampleado em Fez – Being Born (U2, 2009). ATE Normalizado: -4.269 — Causalidade de Granger: 0.85	84
B.9	El Lute / Gotta Go Home (Boney M., 1979) sampleado em Barbra Streisand (Duck Sauce, 2010). ATE Normalizado: 6.555 — Causalidade de Granger: 0.12	85
B.10	Hush, Little Baby (Traditional Folk, 1939) sampleado em Mockingbird (Eminem, 2004). ATE Normalizado: -4.35 — Causalidade de Granger: 0.83	85
B.11	Mo Money Mo Problems (The Notorious B.I.G., 1997) sampleado em Worst Behavior (Drake, 2013). ATE Normalizado: -4.276 — Causalidade de Granger: 0.74	85
B.12	She Will (Lil Wayne, 2011) sampleado em The Motto (Drake, 2011). ATE Normalizado: -4.401 — Causalidade de Granger: 0.02	86
B.13	The Motto (Drake, 2011) sampleado em YOLO (The Lonely Island, 2013). ATE Normalizado: -5.373 — Causalidade de Granger: 1.0	86
B.14	Chaconne in F minor (Johann Pachelbel, 1680) sampleado em What If (Jason Derülo, 2010). ATE Normalizado: -6.801 — Causalidade de Granger: 0.11	86
B.15	Stereo Hearts (Gym Class Heroes, 2011) sampleado em IDGAFOS (Dillon Francis, 2011). ATE Normalizado: 4.832 — Causalidade de Granger: 0.02	87
B.16	Werewolves of London (Warren Zevon, 1978) sampleado em All Summer Long (Kid Rock, 2007). ATE Normalizado: 5.281 — Causalidade de Granger: 0.0	87
B.17	Hide and Seek (Imogen Heap, 2005) sampleado em Whatcha Say (Jason Derülo, 2009). ATE Normalizado: 5.037 — Causalidade de Granger: 0.96	87
B.18	Just a Friend (Biz Markie, 1989) sampleado em Say You're Just a Friend (Austin Mahone, 2012). ATE Normalizado: 3.803 — Causalidade de Granger: 0.0	88
B.19	One Step at a Time (Jordin Sparks, 2007) sampleado em All I See (Kylie Minogue, 2007). ATE Normalizado: 5.98 — Causalidade de Granger: 0.11	88
B.20	Close to Me (The Cure, 1985) sampleado em So Human (Lady Sovereign, 2009). ATE Normalizado: 4.092 — Causalidade de Granger: 0.07	88
B.21	Don't Stand So Close to Me (The Cure, 1985) sampleado em So Human (Lady Sovereign, 2009). ATE Normalizado: -5.471 — Causalidade de Granger: 0.54	89
B.22	Shots (LMFAO, 2009) sampleado em Somebody (Natalie La Rose, 2014). ATE Normalizado: 3.862 — Causalidade de Granger: 0.12	89
B.23	Pon de Floor (Major Lazer, 2009) sampleado em Ass on the Floor (Diddy - Dirty Money, 2010). ATE Normalizado: 5.45 — Causalidade de Granger: 0.02	89

B.24 Dirty Cash (Money Talks) (Adventures of Stevie V, 1989) sampleado em Dirty Cash (Dizzee Rascal, 2009). ATE Normalizado: 10.157 — Causalidade de Granger: 0.34	90
B.25 (I Just) Died in Your Arms (Cutting Crew, 1986) sampleado em Relax, Take It Easy (Mika, 2006). ATE Normalizado: -4.186 — Causalidade de Granger: 0.3	90
B.26 Take It Off (Kesha, 2010) sampleado em Pop Culture (Madeon, 2011). ATE Normalizado: -4.337 — Causalidade de Granger: 0.0	90
B.27 One Way or Another (Blondie, 1978) regravado em One Way or Another (Teenage Kicks) (One Direction, 2013). ATE Normalizado: 5.06 — Causalidade de Granger: 0.93	91
B.28 Tom Ford (Jay-Z, 2013) sampleado em Dirty Vibe (Skrillex, 2014). ATE Normalizado: -4.23 — Causalidade de Granger: 0.48	91
B.29 Valerie (Steve Winwood, 1982) sampleado em Call on Me (Eric Prydz, 2004). ATE Normalizado: 5.351 — Causalidade de Granger: 0.6	91
B.30 I Won't Back Down (Tom Petty, 1989) sampleado em Stay with Me (Sam Smith, 2014). ATE Normalizado: 7.852 — Causalidade de Granger: 0.16	92
B.31 Rocketeer (Far East Movement, 2010) sampleado em Somebody To Love (Justin Bieber, 2010). ATE Normalizado: -42.537 — Causalidade de Granger: 0.0	92
B.32 Rocketeer (Far East Movement, 2010) sampleado em Like a G6 (Far East Movement, 2010). ATE Normalizado: -43.251 — Causalidade de Granger: 0.02	92
B.33 N.Y. State of Mind (Nas, 1994) sampleado em Too Many Rappers (Beastie Boys, 2009). ATE Normalizado: 8.657 — Causalidade de Granger: 0.87	93
B.34 The Motto (Drake, 2011) sampleado em She Will (Lil Wayne, 2011). ATE Normalizado: -10.413 — Causalidade de Granger: 0.03	93
B.35 Baby Got Back (Sir Mix-a-Lot, 1992) sampleado em Ew! (Jimmy Fallon, 2014). ATE Normalizado: -3.575 — Causalidade de Granger: 0.96	93
B.36 Keep on Truckin' (Eddie Kendricks, 1973) sampleado em Do It Well (Jennifer Lopez, 2007). ATE Normalizado: 8.414 — Causalidade de Granger: 0.24	94
B.37 Girls Can't Do What the Guys Do (Betty Wright, 1968) sampleado em Upgrade U (Beyoncé, 2006). ATE Normalizado: -4.615 — Causalidade de Granger: 0.95	94
B.38 Something's Got a Hold on Me (Etta James, 1962) sampleado em Good Feeling (Flo Rida, 2011). ATE Normalizado: 42.553 — Causalidade de Granger: 0.86	94
B.39 Pony (Ginuwine, 1996) sampleado em Jump (Rihanna, 2012). ATE Normalizado: 3.475 — Causalidade de Granger: 0.71	95
B.40 I Will Survive (Gloria Gaynor, 1978) sampleado em "Hush Hush" (Pussycat Dolls, 2008). ATE Normalizado: 3.108 — Causalidade de Granger: 0.74	95

B.41	Heartache Avenue (The Maisonettes, 1983) sampleado em The Avenue (Roll Deep, 2005). ATE Normalizado: -4.585 — Causalidade de Granger: 0.15	95
B.42	Fade to Grey (Visage, 1980) sampleado em One Word (Kelly Osbourne, 2005). ATE Normalizado: 6.286 — Causalidade de Granger: 0.54	96
B.43	Toccatina and Fugue in D minor, BWV 565 (Johann Sebastian Bach, 1707) sampleado em Baby's Coming Back/Transylvania (McFly, 2007). ATE Normalizado: -3.19 — Causalidade de Granger: 0.73	96
B.44	One Nation Under a Groove (Funkadelic, 1978) sampleado em Peaches N Cream (Snoop Dogg song) (Snoop Dogg, 2015). ATE Normalizado: 39.051 — Causalidade de Granger: 0.88	96
B.45	Wanna Be Startin' Somethin' (Michael Jackson, 1982) sampleado em Don't Stop the Music (Rihanna, 2007). ATE Normalizado: 6.412 — Causalidade de Granger: 0.36	97
B.46	No Woman, No Cry (Bob Marley, 1974) sampleado em Rayon de soleil (William Baldé, 2008). ATE Normalizado: -3.275 — Causalidade de Granger: 0.16	97
B.47	Touch It (Busta Rhymes, 2005) sampleado em Technologic (Daft Punk, 2005). ATE Normalizado: -7.117 — Causalidade de Granger: 0.25	97
B.48	Turn Down for What (DJ Snake, 2013) sampleado em Dynamite (Afrojack, 2014). ATE Normalizado: 5.059 — Causalidade de Granger: 1.0	98
B.49	Somebody To Love (Justin Bieber, 2010) sampleado em Rocketeer (Far East Movement, 2010). ATE Normalizado: -4.615 — Causalidade de Granger: 0.19	98
B.50	Like a G6 (Far East Movement, 2010) sampleado em Rocketeer (Far East Movement, 2010). ATE Normalizado: 5.194 — Causalidade de Granger: 1.0	98
B.51	The Rhythm of the Night (Corona, 1993) regravado em Of the Night (song) (Bastille, 2012). ATE Normalizado: 5.556 — Causalidade de Granger: 0.98	99
B.52	In the Closet (Michael Jackson, 1991) sampleado em Take You Down (Chris Brown, 2007). ATE Normalizado: -4.436 — Causalidade de Granger: 0.06	99
B.53	Be My Lover (La Bouche, 1995) sampleado em Be My Lover (Inna, 2013). ATE Normalizado: 4.073 — Causalidade de Granger: 0.51	99
B.54	Leck mich im Arsch (Wolfgang Amadeus Mozart, 1782) sampleado em Leck mich im Arsch (Insane Clown Posse, 2011). ATE Normalizado: 5.039 — Causalidade de Granger: 0.0	100
B.55	Million Voices (Otto Knows, 2012) sampleado em 4 AM (Scooter, 2012). ATE Normalizado: 5.414 — Causalidade de Granger: 0.74	100
B.56	Miss Independent (Ne-Yo, 2008) sampleado em Knock You Down (Keri Hilson, 2009). ATE Normalizado: -4.158 — Causalidade de Granger: 0.93	100
B.57	Like a G6 (Far East Movement, 2010) sampleado em Booty Bounce (Dev, 2010). ATE Normalizado: 6.888 — Causalidade de Granger: 0.99	101

B.58	Nightshift (Commodores, 1985) sampleado em Lose to Win (Fantasia, 2012). ATE Normalizado: -38.059 — Causalidade de Granger: 0.17	101
B.59	Inseparable (Natalie Cole, 1975) sampleado em Karma (Lloyd Banks, 2004). ATE Normalizado: -40.092 — Causalidade de Granger: 0.0	101
B.60	Mr. Lonely (Bobby Vinton, 1964) sampleado em Lonely (Akon, 2004). ATE Normalizado: 5.941 — Causalidade de Granger: 0.72	102
B.61	Party & Bullshit (The Notorious B.I.G., 1993) sampleado em How We Do (Party) (Rita Ora, 2012). ATE Normalizado: 5.901 — Causalidade de Gran- ger: 0.12	102
B.62	Stronger (Britney Spears, 2000) sampleado em Pound the Alarm (Nicki Minaj, 2012). ATE Normalizado: -3.067 — Causalidade de Granger: 0.04	102
B.63	Na Na Hey Hey Kiss Him Goodbye (Steam, 1969) sampleado em Goodbye (Kristinia DeBarge, 2009). ATE Normalizado: 5.522 — Causalidade de Gran- ger: 0.94	103
B.64	My Cherie Amour (Stevie Wonder, 1969) sampleado em Wish You Were Mine (Philip George, 2014). ATE Normalizado: 4.448 — Causalidade de Granger: 0.01	103
B.65	Baba O'Riley (The Who, 1971) sampleado em Escape Velocity (The Chemical Brothers, 2010). ATE Normalizado: 3.984 — Causalidade de Granger: 0.13 .	103
B.66	You Spin Me Round (Like a Record) (Dead or Alive, 1984) sampleado em Right Round (Flo Rida, 2009). ATE Normalizado: 5.577 — Causalidade de Granger: 0.28	104
B.67	Summertime Sadness (Lana Del Rey, 2012) sampleado em Body Electric (Lana Del Rey, 2012). ATE Normalizado: -5.032 — Causalidade de Granger: 0.47 .	104
B.68	Fader (The Temper Trap, 2009) remixado em Fade (Jakwob song) (Jakwob, 2009). ATE Normalizado: -4.615 — Causalidade de Granger: 0.51	104
B.69	Wearing My Rolex (Wiley, 2008) sampleado em Rolex Sweep (Skeptak, 2008). ATE Normalizado: -4.432 — Causalidade de Granger: 0.99	105
B.70	Calabria (Rune, 2003) sampleado em Destination Calabria (Alex Gaudino, 2006). ATE Normalizado: 6.189 — Causalidade de Granger: 0.0	105
B.71	Homecoming (Green Day, 2004) sampleado em American Idiot (Green Day, 2004). ATE Normalizado: -8.073 — Causalidade de Granger: 0.74	105
B.72	We Don't Have to Take Our Clothes Off (Jermaine Stewart, 1986) sampleado em Clothes Off!! (Gym Class Heroes, 2007). ATE Normalizado: 5.624 — Causalidade de Granger: 0.95	106
B.73	Again (Janet Jackson, 1993) sampleado em Solo (Iyaz, 2010). ATE Normali- zado: 4.785 — Causalidade de Granger: 0.14	106

B.74 Let It Rock (Kevin Rudolf, 2008) sampleado em I Made It (Cash Money Heroes) (Kevin Rudolf, 2010). ATE Normalizado: -5.963 — Causalidade de Granger: 0.07	106
B.75 Down (Jay Sean, 2009) sampleado em I Made It (Cash Money Heroes) (Kevin Rudolf, 2010). ATE Normalizado: -3.889 — Causalidade de Granger: 0.41	107
B.76 Ima Korean (Rucka Rucka Ali, 2009) regravado em I Gotta Feeling (Black Eyed Peas, 2009). ATE Normalizado: 7.692 — Causalidade de Granger: 0.02	107
B.77 Heaven Is a Place on Earth (Belinda Carlisle, 1987) regravado em Heaven Is a Place on Earth (Wildside, 2009). ATE Normalizado: 4.68 — Causalidade de Granger: 0.0	107
B.78 She's on Fire (Amy Holland, 1983) sampleado em On Fire (Lil Wayne, 2009). ATE Normalizado: -6.867 — Causalidade de Granger: 0.58	108
B.79 The One That Got Away (Katy Perry, 2010) regravado em The One That Got Away (Jake Coco, 2011). ATE Normalizado: -4.262 — Causalidade de Granger: 0.89	108
B.80 Been Down So Long (The Doors, 1971) sampleado em Hypest Hype (Chase & Status, 2010). ATE Normalizado: 41.03 — Causalidade de Granger: 0.93	108
B.81 Be Happy (Mary J. Blige, 1994) sampleado em Holding You Down (Goin' in Circles) (Jazmine Sullivan, 2010). ATE Normalizado: -4.617 — Causalidade de Granger: 0.1	109
B.82 Music Sounds Better with You (Stardust, 1998) sampleado em Music Sounds Better with U (Big Time Rush, 2011). ATE Normalizado: 3.422 — Causalidade de Granger: 0.9	109

Lista de Tabelas

4.1	Tipos de empréstimos musicais e suas ocorrências no conjunto de dados. . .	39
5.1	Exemplos de músicas próximas ao limiar de corte de 25% para o YouTube. A tabela apresenta títulos originais e seus correspondentes analisados, destacando casos incluídos no conjunto final, apesar de diferenças textuais significativas.	46
5.2	Distribuição da diferença percentual entre visualizações de músicas originais e derivadas por tipo de empréstimo.	51
5.3	Distribuição da diferença percentual entre popularidade de músicas originais e derivadas, baseado na métrica de popularidade do Spotify.	55
6.1	Classes relacionadas a obras musicais no Wikidata, utilizadas na filtragem da coleta. Fonte: https://www.wikidata.org/	61
A.1	Informações disponíveis nos dados provenientes do Youtube	81

Sumário

1	Introdução	19
1.1	Objetivo	22
1.2	Perguntas de pesquisa	22
1.3	Organização do Trabalho	22
2	Trabalhos Relacionados	24
2.1	Empréstimos Musicais	24
2.2	Influência Musical	25
2.3	Mercado de Música Digital e Diversidade no Consumo de Música	26
2.4	Conclusão do Capítulo	27
3	Fundamentos Teóricos	28
3.1	Análise de Séries Temporais	28
3.2	Similaridade entre séries temporais e funções de similaridade	29
3.2.1	Correlação de Pearson	29
3.2.2	Distância Euclidiana	31
3.2.3	Dynamic Time Warping (DTW)	32
3.3	Causalidade de Granger	32
3.4	Regression Discontinuity Design (RDD)	34
4	Conjunto de Dados	37
4.1	WhoSampled	38
4.2	Wikidata	39
4.3	Google Trends	40
4.4	Spotify	42
4.5	Youtube	42
5	RQ1 - Impacto em plataformas de streaming, uma visão estática	44
5.1	Metodologia	45
5.1.1	Coleta de Dados	45
5.1.2	Limpeza e preparação dos dados	45
5.2	Resultados	47
5.2.1	Youtube	47
5.2.2	Spotify	51

5.3	Conclusão	55
6	RQ2 - Empréstimos Musicais e o Interesse pelo Original	57
6.1	Coleta de Dados	57
6.1.1	Identificação das músicas no Google Trends	58
6.1.2	Bases de Conhecimento	59
6.1.3	Coleta e tratamento de informações sobre músicas no Wikidata	59
6.1.4	Coleta do interesse ao longo do tempo no Google Trends	63
6.2	Similaridade entre as Séries de Interesse de Busca	64
6.2.1	Correlação de Pearson	65
6.2.2	Distância Euclidiana	65
6.2.3	Dynamic Time Warping (DTW)	66
6.2.4	Conclusões	67
6.3	Impacto dos samples no interesse pela música original	67
6.3.1	Impacto no curto prazo	67
6.3.1.1	Impacto positivo	69
6.3.1.2	Impacto negativo	70
6.3.2	Impacto no Longo Prazo	72
6.4	Conclusões	74
7	Conclusão e Trabalhos Futuros	75
	Referências	78
	Apêndice A Resumo dos dados obtidos no Youtube	81
	Apêndice B Interesse de Busca vs. RDD Predito	82

Capítulo 1

Introdução

A digitalização transformou drasticamente a produção, distribuição e consumo de música. Tecnologias avançadas de armazenamento, transmissão e processamento de dados proporcionaram mudanças importantes na indústria musical e tornaram a música digital predominante. Este contexto impulsionou novas formas de composição, uma delas conhecida como “empréstimos musicais” (musical borrowings), na qual compositores e produtores exploram influências musicais existentes para conceber novas obras, expandindo assim as fronteiras dos gêneros musicais. Neste trabalho, abordamos três principais formas de empréstimos musicais: amostragem, remix e covers.

A amostragem, também conhecida como *sampling*, é uma técnica que ganhou popularidade no final dos anos 1970. Amostras musicais são partes de músicas pré-existentes que os artistas incorporam em suas próprias composições. Amplamente utilizada em gêneros como hip-hop, eletrônica e música experimental, a amostragem permite a criação de texturas sonoras ¹ únicas, ao combinar cuidadosamente partes de diferentes faixas de maneiras inovadoras. Além disso, a manipulação das amostras através de processos como cortes (*slicing*), ajustes de tom (*pitching*) e alongamento temporal (*time-stretching*) oferece possibilidades infinitas para a criação de novos arranjos e melodias.

Os remixes representam também uma importante estratégia de composição musical. Ao criar um remix, os artistas utilizam uma música existente e a transformam por meio de alterações no arranjo, estrutura, ritmo, textura e efeitos sonoros. Empregados em gêneros como eletrônica, dance e pop, os remixes revitalizam músicas, tornando-as mais modernas e atraentes para as pistas de dança. Além disso, os remixes permitem explorar novas direções estilísticas e colaborar com outros artistas para alcançar diferentes audiências.

Covers também são uma prática comum na indústria da música e ocorrem quando um artista grava sua própria versão de uma música previamente lançada por outro artista. Essa prática permite que novos artistas ganhem visibilidade ao reinterpretar sucessos já conhecidos, ao mesmo tempo em que prestam homenagem ao trabalho original. As interpretações podem variar, sendo desde as que são quase idênticas à versão original até

¹Citamos “texturas sonoras” não só como a forma com que vários sons se combinam em uma composição, mas também a expressão do pensamento através do som, como abordado em [Bairon \[2005\]](#).

arranjos completamente novos que oferecem uma perspectiva diferente da música. A popularidade dos covers pode muitas vezes ser atribuída ao reconhecimento pré-existente da música original, proporcionando uma base de fãs pronta para receber a nova versão. Além disso, em alguns casos, os covers podem superar a popularidade das versões originais, especialmente quando o novo artista traz um estilo ou uma abordagem única que ressoa com o público contemporâneo.

O fácil acesso a equipamentos de gravação e produção e o surgimento de novas plataformas de distribuição, como os serviços de streaming, tornaram o arranjo da indústria musical mais descentralizado. Em um passado recente, grandes corporações da indústria, como gravadoras e selos, detinham o controle sobre as estratégias de produção e lançamento das músicas. Agora, produções independentes passaram a ter mais espaço, tornando a produção musical mais democrática e dando aos artistas maior controle sobre elas. Nesse sentido, a colaboração entre artistas tem sido uma estratégia utilizada para ampliar o público e manter o sucesso [Seekhao \[2020\]](#). Empréstimos musicais fazem parte dessa estratégia, pois representam formas de colaboração em que o compositor utiliza elementos de outras obras, mesmo que sem o consentimento do autor original. Esse tipo de composição é intencional e evidencia a influência de obras sobre outras, uma vez que o compositor seleciona deliberadamente uma obra em particular.

A música “Crazy in Love”, da artista pop norte-americana Beyoncé, lançada em 2003 e considerada uma de suas músicas mais icônicas, é um exemplo de empréstimo musical. Um dos elementos que mais chamaram atenção foi a introdução da música, marcada pela entrada triunfal dos metais. De acordo com o blog [Hypeness \[2020\]](#), esses metais foram extraídos da música “Are You My Woman” do quarteto americano The Chi-Lites, um sucesso dos anos 1970 (veja Figura 1.1). Embora o sample tenha sido retirado de uma música criada em uma era pré-digital, podemos compreender os efeitos desse sample na obra de Beyoncé, embora seja difícil medir sua contribuição em relação ao sucesso da música.

O estudo das dinâmicas dos empréstimos musicais também está ligado às questões jurídicas, já que existem legislações que regulamentam o uso de obras autorais. Ao compreender os efeitos do empréstimo musical nas músicas envolvidas, pode-se contribuir para novos entendimentos no campo do direito autoral, bem como nas decisões de músicos, empresários e do mercado musical como um todo. Por exemplo, o estudo intitulado “Sampling Increases Music Sales: An Empirical Copyright Study” destaca que, nos Estados Unidos, a Suprema Corte compreende que o fator determinante para a necessidade de autorização para um empréstimo musical (uso justo) é o impacto do empréstimo no mercado da obra original [Schuster et al. \[2019\]](#). Os resultados do estudo indicam que a amostragem digital aumenta as vendas das músicas amostradas (música original), sugerindo que o uso de samples pode ser considerado uso justo. O estudo conclui que o mercado atual de licenciamento de samples é resultado de uma jurisprudência desfavorável à amostragem,

que pode ser resultado da falta de dados empíricos relevantes.

No contexto brasileiro, a Lei de Direitos Autorais (Lei n.º 9.610/1998) não prevê exceções semelhantes ao uso justo (*fair use*), adotado em sistemas de *common law*, como o norte-americano Bra [1998]. A legislação brasileira exige autorização expressa do titular dos direitos autorais para a utilização de trechos de obras musicais, ainda que em pequena proporção, o que torna o cenário nacional mais restritivo em relação à prática de *sampling*. Tal situação aproxima o Brasil de outras jurisdições de tradição romano-germânica, em que a proteção patrimonial e moral do autor tende a prevalecer sobre flexibilizações de mercado.

The screenshot shows the WhoSampled website interface. At the top, there is a navigation bar with links for 'News', 'Discover', 'Sample Packs', 'Submit', and '6D'. A search bar is located on the right side of the navigation bar. Below the navigation bar, the main content area displays the track 'Crazy in Love' by Beyoncé feat. Jay-Z. The track information includes the album 'Dangerously in Love', the year '2003', and the producers 'Rich Harrison, Beyoncé'. There is a 'Buy this Track' button. Below the track information, there is a section for 'Song Connections'. This section is divided into two parts: 'Contains samples of 1 song' and 'Sampled in 43 songs'. The 'Contains samples of 1 song' section shows a connection to 'Are You My Woman? (Tell Me So)' by Chi-Lites (1970), with a label 'Multiple Elements' and sub-labels 'Soul / Funk / Disco'. The 'Sampled in 43 songs' section shows three connections: 'Booty' by Saucy Santana feat. Latto (2022) with a label 'Multiple Elements' and sub-labels 'Hip-Hop / Rap / R&B'; 'Pump It Up' by Girl Talk (2003) with a label 'Hook / Riff' and sub-label 'Other'; and 'Get Me Bodied (Extended Mix)' by Beyoncé (2006) with a label 'Vocals / Lyrics' and sub-labels 'Rock / Pop'. A 'see all' button is located at the end of the 'Sampled in 43 songs' section.

Figura 1.1: Site WhoSampled.com exibindo informações sobre conexões musicais da obra “Crazy in Love” de Beyoncé. Fonte: WhoSampled.com

Compreender as consequências do uso de samples nas músicas originais pode contribuir para discussões mais justas no âmbito jurídico e decisões mais embasadas por parte de músicos e empresários. Dessa forma, os resultados desta pesquisa podem beneficiar a indústria musical, otimizando receitas por meio de decisões mais informadas sobre a produção musical baseada em empréstimos, além de fornecer fundamentos para a revisão de legislações sobre direitos autorais.

1.1 Objetivo

Partindo da hipótese de que a prática de amostragem aumenta a popularidade das músicas originais, o objetivo central deste trabalho é investigar o impacto dos empréstimos musicais, como samples, covers e remixes, na popularidade das músicas envolvidas. Especificamente, buscamos compreender como esses tipos de empréstimos influenciam a popularidade, tanto da música que os utiliza quanto daquela que os fornece.

1.2 Perguntas de pesquisa

Para alcançar o objetivo central e testar a hipótese de pesquisa, este trabalho busca responder às seguintes perguntas:

RQ1: Como a amostragem musical impacta estaticamente a popularidade das músicas originais em plataformas de streaming, como Spotify e YouTube?

RQ2: Como a amostragem musical influencia dinamicamente o interesse de busca pelas músicas originais ao longo do tempo?

1.3 Organização do Trabalho

Este trabalho está estruturado em 7 capítulos, incluindo esta Introdução, conforme detalhado a seguir:

- No Capítulo 2, é realizada uma revisão da literatura sobre empréstimos musicais, análise de influência musical e o mercado de música digital. São explorados estudos que investigam os impactos da digitalização, a prática de empréstimos musicais, a influência entre músicas e artistas, além da diversidade no consumo musical. Essa revisão serve como base teórica para o desenvolvimento das análises subsequentes.
- O Capítulo 3 apresenta o referencial teórico adotado para fundamentar as análises realizadas e as conclusões discutidas nos capítulos seguintes.

-
- No Capítulo 4, são descritos os conjuntos de dados utilizados neste estudo, detalhando suas características e fontes.
 - O Capítulo 5 realiza uma análise comparativa das métricas de sucesso entre músicas originais e suas versões sampleadas, covers e remixadas, utilizando dados estáticos do YouTube e Spotify.
 - No Capítulo 6, é explorada a técnica Regression Discontinuity Design (RDD) para estimar o efeito causal do lançamento de um sample no interesse pela música original, utilizando dados de séries temporais de buscas obtidos no Google Trends. Além disso, aplica-se a Causalidade de Granger para avaliar também o impacto de longo prazo.
 - O Capítulo 7 apresenta as conclusões deste estudo, sintetizando os principais resultados, limitações e discutindo as contribuições para a área de estudos musicais e a indústria da música. Também são propostas direções para pesquisas futuras com base nos achados deste trabalho.

Capítulo 2

Trabalhos Relacionados

Estudos já investigaram os impactos da digitalização da música na indústria como um todo, bem como a influência de técnicas de sample, remix e cover na colaboração musical e no sucesso comercial dos artistas. Nas próximas seções, abordaremos alguns desses estudos e discutiremos suas contribuições para a compreensão dessas dinâmicas.

2.1 Empréstimos Musicais

O termo “empréstimos musicais” (do inglês, *Musical Borrowings*) é definido como o estudo dos diferentes procedimentos que compositores utilizam para criar suas obras baseando-se em melodias, planos tonais ou estruturas pré-existentes [Burkholder \[1994\]](#). Este tema tem sido estudado não apenas no campo da música, mas também no direito, especialmente no que se refere aos direitos autorais.

Muitos autores discutem o processo de composição e como a autoria de uma música é frequentemente interpretada pela legislação de direitos autorais de maneira inadequada. Segundo [Arewa](#), “a doutrina dos direitos autorais incorpora noções de autoria romantizada que assumem criação independente e autônoma, e até mesmo genial, de obras musicais originais”¹ [Arewa \[2005\]](#). No mesmo trabalho, o autor exemplifica a discrepância entre a teoria jurídica sobre o assunto e a prática musical, tomando como caso a música hip-hop, o segundo gênero musical mais popular nos Estados Unidos e força musical e cultural significativa em todo o mundo. Ele menciona que o advento do hip-hop levantou sérias preocupações com a lei de direitos autorais devido ao uso de técnicas como o *sampling*. Os tribunais norte-americanos consideraram o *sampling* violação dos direitos

¹A citação original em [Arewa \[2005\]](#) destaca que o empréstimo musical é um aspecto onipresente na criação musical de todos os gêneros e períodos. A doutrina dos direitos autorais não reflete adequadamente essa realidade, pois incorpora noções de autoria romântica que pressupõem autoria independente e autônoma, enfatizando uma visão individualista e autônoma da autoria musical que é central para a lei de direitos autorais, desconsiderando a importância e a continuidade das práticas de empréstimo musical.

autorais, e o autor pondera que o uso recorrente de técnicas de empréstimos musicais requer consideração cuidadosa, visto que a cópia e o empréstimo têm sido e podem continuar a ser fontes de inovação na música.

Em contrapartida, as práticas de empréstimos musicais podem afetar grupos vulneráveis e aumentar desigualdades. Hesmondhalgh [2006] explora essa prática ao observar o uso, por músicos brancos, de obras criadas por músicos negros, buscando entender como esse tipo de prática pode fortalecer formas sistêmicas de desigualdade social e cultural. De maneira similar, podemos refletir sobre a produção musical em mercados não hegemônicos, como nas regiões da América Latina e África, onde tais práticas também podem ter implicações significativas.

2.2 Influência Musical

A análise da influência musical é amplamente reconhecida na literatura como uma tarefa complexa, devido à sua subjetividade e às diversas formas pelas quais os compositores podem ser influenciados. Em análises baseadas em dados, os desafios aumentam, pois pode ser inviável modelar essas influências de maneira precisa.

Shalit et al. [2013] abordaram a análise da influência musical aplicando técnicas de modelagem de tópicos (*Dynamic Topic Model*). Para isso, o autor utilizou um conjunto de dados contendo 24.941 músicas e 9.222 artistas, abrangendo o período de 1922 a 2010. O modelo desenvolvido foi utilizado para entender propriedades da influência, sendo treinado com características das músicas, como timbre, tempo, tonalidade (maior ou menor), intensidade, entre outras. As variáveis resposta são as pontuações para cada tópico de influência e as atribuições agregadas de tópico-palavra. Essas variáveis definem a mistura de tópicos de uma música e o quanto ela influencia cada um dos tópicos. Os autores descobriram que a influência musical e a inovação musical não são monotonicamente correlacionadas. Além disso, identificaram que canções mais influentes foram mais inovadoras durante dois períodos: o início dos anos 1970 e meados dos anos 1990.

Bryan and Wang [2011] exploram redes de influência e ranqueamento em músicas baseadas em *sampling*, utilizando um conjunto de dados do WhoSampled.com com 42.447 registros de amostragens musicais. O estudo investiga o uso de *samples* em diferentes gêneros, identificando instrumentos frequentemente reutilizados, como bateria, baixo e vozes, e aplica técnicas de análise de redes para medir a influência de músicas, artistas e gêneros. A abordagem introduzida, chamada *Collapse-and-Sum*, unifica redes de influência em níveis hierárquicos, permitindo derivar matrizes de influência para artistas e gêneros a partir de uma matriz inicial baseada em músicas, com um fator de decaimento

que controla a propagação indireta da influência. Além disso, o artigo destaca tendências temporais e por gênero, revelando, por exemplo, a predominância de *samples* de funk, soul e disco em músicas de hip-hop e eletrônica. Apesar de apresentar importantes padrões musicológicos, o trabalho não investiga diretamente a relação entre o uso de *samples* e a popularidade comercial das músicas, o que limita sua aplicação em análises de impacto no mercado musical.

Em Ortega [2021], o autor analisa a produção de *covers* para medir o impacto musical de uma obra entre os artistas. Para analisar esse impacto, foi construída uma rede com mais de 106 mil artistas e 855 mil *covers* extraídas do site Second-Hand Songs². O objetivo do autor com essa rede é entender como as versões *covers* podem informar sobre a história da música popular contemporânea. Foram utilizadas várias métricas para identificar artistas mais influentes, transversais e com mais regravações. Entre os principais achados, a pesquisa mostrou que o gênero é o principal critério em uma colaboração, seguido pelo idioma dos artistas.

2.3 Mercado de Música Digital e Diversidade no Consumo de Música

A revolução da música digital transformou profundamente não apenas a forma como a música é consumida e criada, mas também os modelos de negócios que a sustentam. Em seu estudo sobre o mercado musical, Bello and Garcia [2021] analisa essas mudanças, destacando que, nos dias atuais, “a música é distribuída a um custo marginal nulo, o que significa que uma faixa de áudio pode ser reproduzida infinitas vezes sem nenhum custo extra”. O autor também explora como as plataformas de streaming superaram as limitações físicas das lojas tradicionais ao definir preços e oferecer serviços, “tornando a oferta musical mais diversificada”.

O estudo ainda aponta que “há evidências de que a maior disponibilidade de música foi acompanhada por maior diversidade e quantidade de consumo musical”. Analisando a evolução dessa diversidade ao longo dos anos, o autor identifica tendências globais no mercado musical, baseando-se em dados abrangentes do Spotify e do iTunes, cobrindo quatro anos e 39 países. Os resultados indicam um aumento significativo no número de músicas, artistas e gravadoras presentes nas paradas em comparação a alguns anos atrás, resultando em rankings nacionais mais diversos ao redor do mundo.

Essa transformação é interpretada como um processo de divergência cultural, no

²<https://secondhandsongs.com/>

qual os países tornam-se cada vez mais distintos em relação às músicas que dominam suas paradas. Por esse motivo, estudos que consideram exclusivamente rankings globais, como o *Billboard*, podem apresentar limitações ao desconsiderar a diversidade musical presente em cada país.

2.4 Conclusão do Capítulo

A revisão de trabalhos relacionados evidenciou que, embora exista um corpo teórico consolidado sobre autoria, redes de influência e transformação do consumo musical, ainda permanecem lacunas importantes quanto à mensuração empírica dos impactos que práticas como sampling, remixes e covers exercem sobre a popularidade das obras originais. A presente pesquisa se insere nesse contexto ao propor uma abordagem que alia dados de plataformas digitais e técnicas de inferência causal, buscando quantificar esses efeitos.

Dessa forma, a dissertação avança no estudo dos efeitos dos empréstimos musicais sobre a popularidade das obras originais, diferenciando-se de trabalhos anteriores ao examinar diretamente a relação entre versões derivadas e sucesso comercial. Entre suas principais contribuições, destacam-se: a análise do impacto comercial; a combinação de abordagens estática e dinâmica com metodologias avançadas, como Regression Discontinuity Design e Causalidade de Granger; o preenchimento de lacunas empíricas importantes para decisões jurídicas sobre amostragem musical; e as implicações práticas para a indústria no que se refere à otimização de receitas e à adaptação a diferentes plataformas digitais. Com isso, o estudo aprofunda a compreensão da dinâmica de popularidade das músicas originais frente aos empréstimos musicais, oferecendo subsídios relevantes para a academia, a indústria e a legislação.

Capítulo 3

Fundamentos Teóricos

Este capítulo apresenta os conceitos teóricos utilizados nas análises realizadas neste trabalho. Discutimos inicialmente séries temporais e métodos de similaridade que permitem avaliar variações entre sequências de dados. Em seguida, introduzimos a causalidade de Granger, uma ferramenta estatística para identificar potenciais relações de causa e efeito entre séries temporais. Por fim, abordamos o *Regression Discontinuity Design* (RDD), um método usado para inferir impactos causais em cenários quase-experimentais, com foco em sua aplicação para avaliar o impacto do lançamento de versões musicais sobre o interesse público.

3.1 Análise de Séries Temporais

A análise de séries temporais é uma ferramenta poderosa com inúmeras aplicações, incluindo a explicação de fenômenos passados, a compreensão dos mecanismos subjacentes, a previsão de comportamentos futuros e a otimização do controle de sistemas.

Séries temporais são conjuntos de observações de uma variável organizadas em uma sequência temporal ordenada e equidistante. Suas aplicações abrangem áreas como finanças, economia, ciências sociais, meteorologia, entre outras.

Formalmente, uma série temporal X é representada como uma sequência de observações $X = \{x_1, x_2, \dots, x_t\}$, onde t denota o tempo. Cada observação x_t é registrada em intervalos de tempo regulares, como dias, meses ou anos. Elas podem ser univariadas, quando uma única variável é medida ao longo do tempo, ou multivariadas, ao observar simultaneamente diversas variáveis.

Essas observações podem ser classificadas como contínuas ou discretas. No caso contínuo, os registros são feitos de forma contínua ao longo do tempo, como em sinais elétricos ou voltagens. Já no caso discreto, as medições ocorrem em intervalos específicos, como taxas de juros, preços de commodities ou, no contexto deste trabalho, o volume de consultas em ferramentas de busca. A Figura 3.1 ilustra uma série temporal discreta que

representa a evolução do interesse de busca por músicas na internet.

Outro aspecto essencial das séries temporais é a dependência e correlação entre suas observações, o que torna a ordem das observações fundamental para a análise [Wei \[2006\]](#). Essa característica apresenta desafios para métodos estatísticos convencionais, que geralmente assumem que as observações são independentes e identicamente distribuídas.

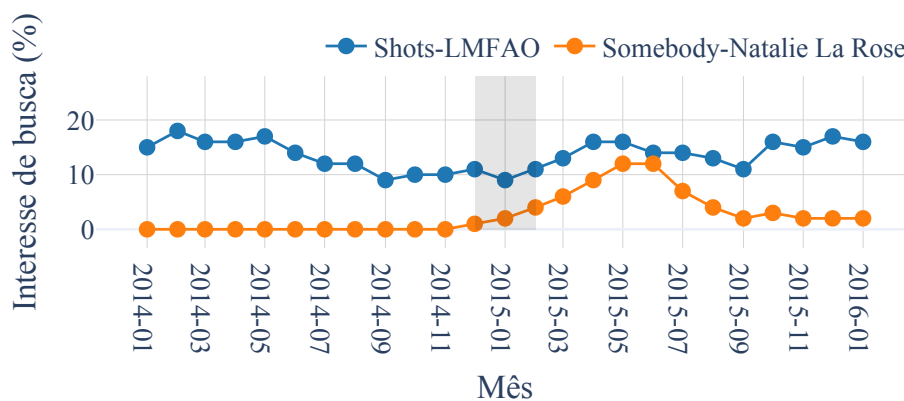


Figura 3.1: Evolução temporal do interesse de busca por músicas

3.2 Similaridade entre séries temporais e funções de similaridade

A avaliação da similaridade entre distintas séries é central na análise de séries temporais e desempenha papel importante em diversas aplicações, abrangendo desde o reconhecimento de padrões até o agrupamento e a detecção de anomalias. A abordagem para essa análise varia conforme o domínio específico e os objetivos particulares em questão. A seguir, delineamos algumas das técnicas empregadas para quantificar a similaridade entre séries temporais.

3.2.1 Correlação de Pearson

A correlação de Pearson é uma medida estatística amplamente utilizada para avaliar a intensidade e a direção da relação linear entre duas variáveis. Em contextos de séries temporais, essa métrica quantifica o grau de associação entre os valores de duas

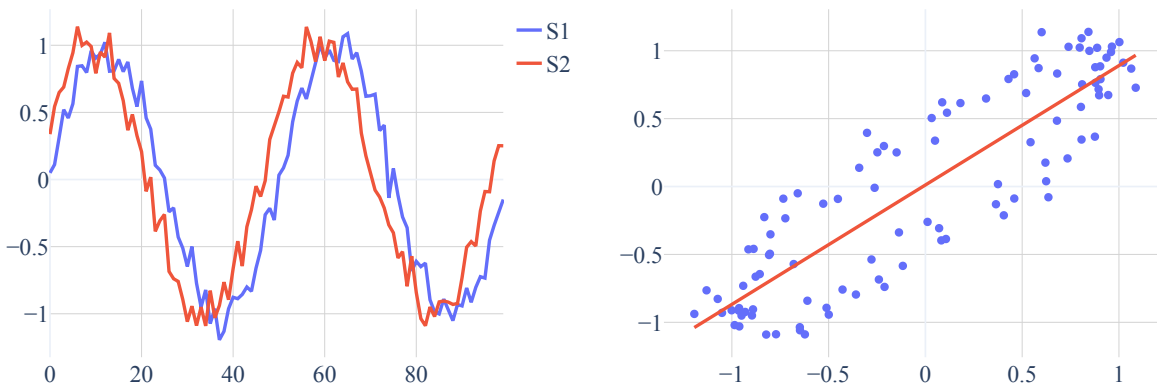
séries ao longo do tempo. A correlação de Pearson varia entre -1 e 1, onde -1 indica uma correlação linear negativa perfeita, 1 representa uma correlação linear positiva perfeita e 0 denota a ausência de correlação linear.

A fórmula que define a correlação de Pearson entre duas séries X e Y é expressa como:

$$\rho_{X,Y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (3.1)$$

x_i e y_i são os valores das séries X e Y , respectivamente, enquanto \bar{x} e \bar{y} representam suas médias. O numerador dessa equação calcula a soma dos produtos das diferenças entre cada valor e sua média correspondente, enquanto o denominador é a raiz quadrada do produto das somas dos quadrados dessas diferenças.

A Figura 3.2 apresenta um exemplo ilustrativo da aplicação da correlação de Pearson em séries temporais. A Figura 3.2a mostra duas séries temporais com uma correlação geral de 0.87, indicando uma relação linear positiva de alta intensidade. Isso pode ser confirmado avaliando a inclinação da curva de tendência na Figura 3.2b.



(a) Sobreposição das séries temporais com correlação de 0.87

(b) Gráfico de dispersão das séries temporais, evidenciando a tendência linear positiva.

Figura 3.2: Exemplo da aplicação da correlação de Pearson em séries temporais, com evidências de forte associação linear positiva.

Apesar de sua utilidade, a correlação de Pearson possui limitações importantes. Primeiramente, ela não captura relações não lineares entre as variáveis, restringindo sua aplicação a contextos onde a linearidade é predominante. Além disso, essa métrica é sensível a outliers, que podem distorcer significativamente os valores calculados. Por fim, a correlação de Pearson pressupõe que os dados sejam homoscedásticos, ou seja, que a variância seja constante ao longo de todo o intervalo de análise [Jin Cheong \[2024\]](#).

3.2.2 Distância Euclidiana

A distância Euclidiana é uma métrica simples que calcula a distância entre os pontos correspondentes de duas séries temporais no espaço Euclidiano. Seja $X = \{x_1, x_2, \dots, x_t\}$ e $Y = \{y_1, y_2, \dots, y_t\}$ duas séries temporais de comprimento t . A distância Euclidiana $d(X, Y)$ entre X e Y é definida como:

$$d(X, Y) = \sqrt{\sum_{i=1}^t (x_i - y_i)^2}. \quad (3.2)$$

Aqui, x_i e y_i representam os valores das séries temporais X e Y no instante i , respectivamente. Quanto menor a distância, maior a similaridade das séries analisadas.

Apesar de ser de fácil compreensão e aplicação, a distância Euclidiana apresenta limitações, pois não leva em consideração a estrutura temporal das séries nem é sensível a deslocamentos e escalas. A Figura 3.3 ilustra essas limitações ao comparar a distância Euclidiana com o DTW. Note que, embora as séries temporais tenham sido deslocadas verticalmente para melhor visualização, os valores das características (eixo y) devem ser considerados equivalentes. Essa comparação evidencia que o DTW é capaz de alinhar as séries temporalmente, mesmo em casos de discrepâncias como atrasos ou diferenças na escala, enquanto a distância Euclidiana mede apenas a diferença ponto a ponto, desconsiderando essas variações. O DTW será detalhado na próxima seção, onde suas vantagens e aplicações serão exploradas.

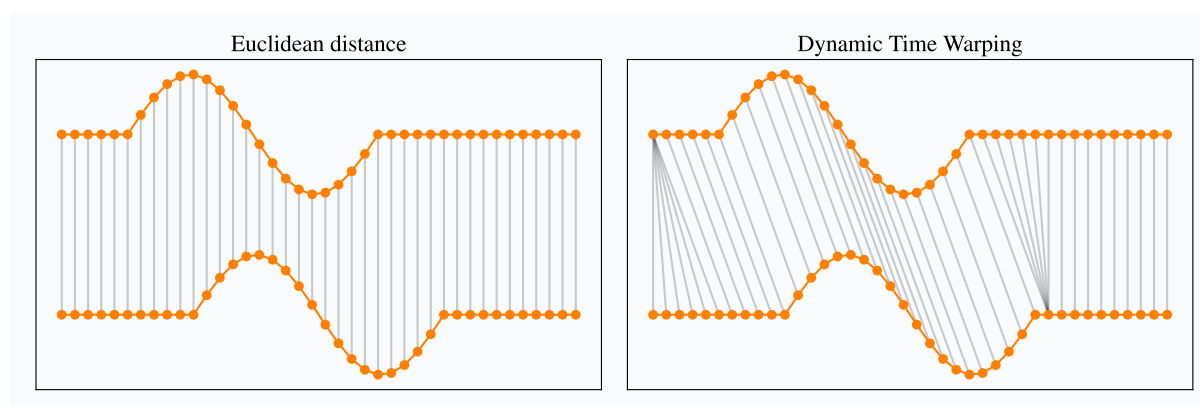


Figura 3.3: Comparação entre DTW e distância euclidiana. Para fins de visualização, as séries temporais estão deslocadas verticalmente. Fonte: [Tavenard \[2021\]](#)

3.2.3 Dynamic Time Warping (DTW)

O *Dynamic Time Warping* (DTW) é uma técnica que calcula a similaridade entre duas séries temporais, alinhando-as de maneira ótima no tempo e permitindo deformações temporais locais. Essa abordagem é robusta a deslocamentos e variações de escala, sendo especialmente útil para séries temporais com variações de velocidade ou distorções temporais.

Sejam X e Y duas séries temporais, o primeiro passo no cálculo da distância DTW é construir a matriz de custo acumulado $C(i, j)$, que representa o custo acumulado de alinhar os prefixos das séries $X(1 : i)$ e $Y(1 : j)$. A relação de recorrência para calcular $C(i, j)$ é dada por:

$$C(i, j) = d(x_i, y_j) + \min(C(i-1, j), C(i, j-1), C(i-1, j-1)) \quad (3.3)$$

para $1 \leq i \leq n$ e $1 \leq j \leq m$, onde $d(x_i, y_j)$ é o custo de correspondência local entre os pontos x_i e y_j das séries X e Y , respectivamente.

Após calcular a matriz de custo acumulado, o próximo passo é construir o *Warping Path*, que representa o caminho de alinhamento ótimo entre as séries X e Y . O *Warping Path* é determinado retrocedendo a partir do ponto final (n, m) na matriz de custo acumulado, seguindo os movimentos diagonais, horizontais e verticais conforme a célula adjacente com o menor custo acumulado. A distância DTW é então calculada como a soma dos custos acumulados ao longo do *Warping Path*. Uma distância DTW igual a 0 indica que as duas séries temporais são perfeitamente semelhantes ou correlacionadas.

Apesar de suas vantagens, o DTW apresenta algumas limitações. A definição de um limiar para determinar se duas séries temporais são semelhantes pode ser subjetiva e depende do contexto da aplicação. Além disso, o DTW pode ser computacionalmente intensivo, especialmente para séries longas, devido à sua complexidade de $O(nm)$, onde n e m são os comprimentos das séries. Outro ponto importante é que o DTW pode ser sensível a ruídos e outliers, o que pode comprometer a precisão da medida de similaridade.

3.3 Causalidade de Granger

A Causalidade de Granger é uma metodologia estatística amplamente utilizada para investigar relações causais entre duas séries temporais, X e Y . Esse conceito baseia-se na ideia de que uma série X é considerada causal para Y se os valores passados de X

fornecerem informações úteis para a predição de Y , além do que é explicado pelos valores passados de Y sozinhos.

O teste de causalidade de Granger envolve a estimativa e comparação de dois modelos de regressão para y_t : um que considera apenas os valores defasados de Y e outro que incorpora, adicionalmente, os valores defasados de X . Esses modelos podem ser representados como:

$$y_t = \alpha + \sum_{i=1}^p \beta_i y_{t-i} + \epsilon_t, \quad (3.4)$$

$$y_t = \alpha' + \sum_{i=1}^p \beta'_i y_{t-i} + \sum_{j=1}^q \gamma'_j x_{t-j} + \epsilon'_t, \quad (3.5)$$

onde:

- α , β_i e ϵ_t são, respectivamente, o intercepto, os coeficientes e o termo de erro do modelo baseado apenas em Y ;
- α' , β'_i , γ'_j e ϵ'_t correspondem aos parâmetros e ao termo de erro do modelo que inclui também X ;
- p e q são os números de defasagens (ou lags) para Y e X , respectivamente.

Para avaliar a significância da inclusão de X , calculam-se as somas dos quadrados dos resíduos (SSR) de ambos os modelos:

$$SSR = \sum_{t=1}^T \epsilon_t^2, \quad SSR' = \sum_{t=1}^T \epsilon'_t{}^2, \quad (3.6)$$

onde SSR se refere ao modelo sem X e SSR' ao modelo que incorpora X .

A estatística F , utilizada para o teste, é dada por:

$$F = \frac{(SSR - SSR')/q}{SSR'/(T - p - q - 1)}, \quad (3.7)$$

em que T é o número total de observações. O valor calculado de F é comparado a um valor crítico da distribuição F . Se F exceder o valor crítico, rejeita-se a hipótese nula H_0 , que afirma a ausência de causalidade de X para Y . Nesse caso, conclui-se que X Granger-cause Y .

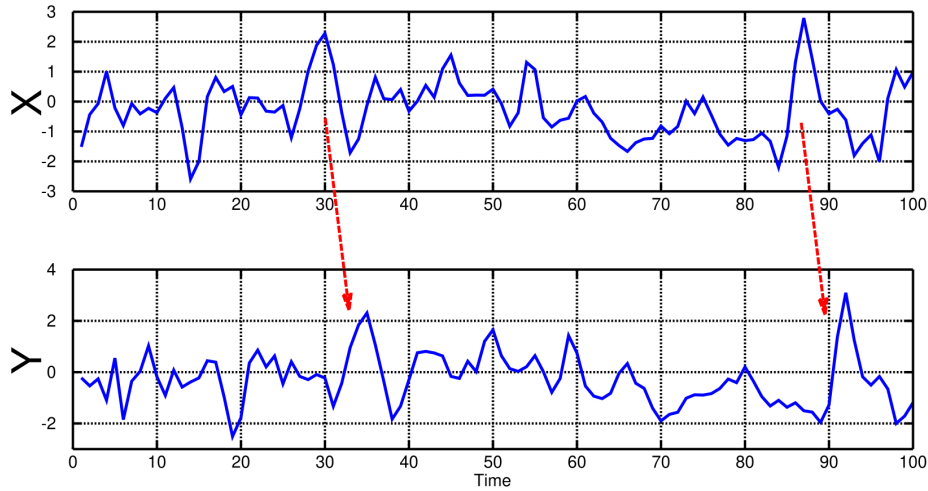


Figura 3.4: Ilustração de causalidade de Granger entre as séries temporais X e Y . Padrões de X são repetidos em Y após um atraso temporal, conforme indicado pelas setas. Fonte: [BiObserver \[2025\]](#).

3.4 Regression Discontinuity Design (RDD)

O *Regression Discontinuity Design* (RDD) é uma abordagem quase-experimental utilizada para avaliar o impacto de uma intervenção ou tratamento quando a atribuição ao grupo de tratamento é determinada por um ponto de corte estrito. Esse método é particularmente útil em cenários onde a randomização não é viável, pois explora a descontinuidade no ponto de corte para fazer inferências sobre efeitos causais.

O princípio fundamental do RDD consiste em comparar os resultados de observações situadas logo acima e abaixo do ponto de corte. A suposição é que as observações em ambos os lados do limite são semelhantes em todos os aspectos, exceto pela exposição ao tratamento. Dessa forma, qualquer mudança abrupta no resultado no ponto de corte pode ser atribuída ao efeito do tratamento. No entanto, é importante destacar que o RDD, por si só, não permite uma inferência causal definitiva, pois não elimina automaticamente a influência de possíveis variáveis de confusão.

Considere uma série temporal $X(t)$, onde t representa o tempo, e um ponto de corte t_c que marca o início do tratamento. Os períodos com $t \geq t_c$ são tratados, enquanto $t < t_c$ não são tratados. O efeito causal do tratamento, denotado por τ , é estimado pela descontinuidade nos valores esperados de $X(t)$ no ponto t_c :

$$\tau = \lim_{t \rightarrow t_c^+} \mathbb{E}[X(t) | t] - \lim_{t \rightarrow t_c^-} \mathbb{E}[X(t) | t]. \quad (3.8)$$

Este efeito é estimado com base na suposição de que, na ausência do tratamento, $X(t)$ seria contínuo no ponto t_c . Portanto, qualquer descontinuidade observada em $X(t)$ no ponto de corte pode ser atribuída ao efeito do tratamento. A Figura 3.5 apresenta um exemplo do uso do RDD, mostrando a relação entre a variável contínua X e o resultado Y , destacando a descontinuidade no ponto de corte c .

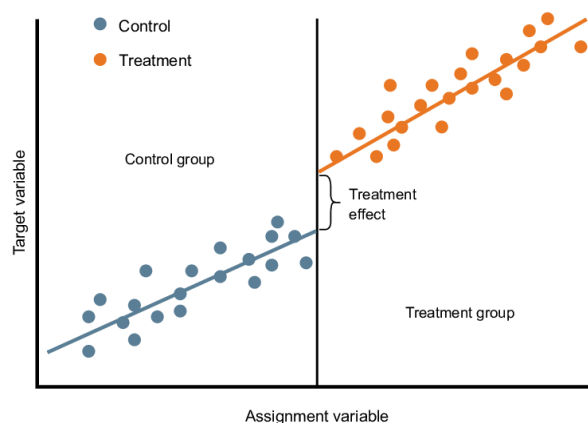


Figura 3.5: Representação gráfica do Regression Discontinuity Design. A figura mostra os grupos de controle e tratamento, destacando o efeito médio do tratamento (ATE) no ponto de corte. Fonte: Liu et al. [2022].

O RDD baseia-se em pressupostos para garantir a validade da estimativa do efeito causal. São eles:

1. Continuidade da variável contínua: A distribuição da variável contínua ao redor do ponto de corte deve ser suave.
2. Ausência de manipulação da variável contínua: Unidades não podem se organizar sistematicamente ao redor do ponto de corte.
3. Randomização local: Unidades próximas ao ponto de corte são assumidas como semelhantes em características observadas e não observadas.

Existem dois principais tipos de RDD, exato e aproximado. No RDD exato, a atribuição ao tratamento é determinística e depende exclusivamente do ponto de corte, ou seja, $D = \mathbb{I}(X \geq c)$, onde D é o indicador de tratamento. Por outro lado, no RDD aproximado, a atribuição ao tratamento é probabilística, sendo influenciada pela variável contínua, mas com a probabilidade de receber o tratamento apresentando uma mudança descontínua no ponto de corte.

Por fim, a aplicação do RDD em estudos de causalidade exige a cuidadosa escolha do ponto de corte e a consideração das características do contexto específico. Embora seja eficaz na avaliação de políticas públicas, mudanças de regulamentação e outros fenômenos onde a randomização não é possível, o RDD apresenta algumas limitações importantes. A

validade das inferências depende da correta especificação do ponto de corte, da ausência de manipulação da variável contínua e de um número suficiente de observações próximas ao limiar. Além disso, os efeitos estimados são localizados, ou seja, aplicáveis principalmente às unidades próximas ao ponto de corte, o que restringe a generalização para toda a população. Compreender essas limitações é essencial para interpretar corretamente os resultados e evitar conclusões causais indevidas.

Capítulo 4

Conjunto de Dados

Este capítulo apresenta os conjuntos de dados utilizados neste estudo e descreve as estratégias de coleta e integração adotadas para sua construção. A Figura 4.1 ilustra o fluxo completo do processo de coleta e integração dos dados.



Figura 4.1: Fluxo do processo de coleta e integração dos dados. A figura destaca as fontes utilizadas e os métodos empregados para responder às perguntas de pesquisa.

Ao investigar o impacto dos empréstimos musicais na popularidade das músicas originais, conforme discutido na Seção 1.2, utilizamos dados da plataforma WhoSampled para identificar os empréstimos musicais. A partir desse mapeamento, complementamos as informações com métricas de popularidade musical, como visualizações, execuções e interesse de busca, obtidas em outras plataformas.

No caso da Pergunta de Pesquisa **RQ1**, que avalia o impacto estático da amostragem musical na popularidade das músicas originais, utilizamos dados do YouTube, considerando o número de visualizações de vídeos musicais, e do Spotify, que fornece métricas consolidadas de popularidade.

Para a Pergunta de Pesquisa **RQ2**, que analisa o impacto dinâmico da amostragem musical, recorreremos inicialmente ao Wikidata para obter identificadores das músicas em bases de conhecimento. Esses identificadores permitiram acessar o histórico de interesse de busca das músicas no Google Trends de forma confiável.

4.1 WhoSampled

O *WhoSampled.com* é uma plataforma que mantém um banco de dados abrangente de músicas baseadas em samples, covers e remixes. Fundado por Nadav Poraz em 2008, o site cataloga empréstimos musicais com dados fornecidos por contribuidores, moderadores e a própria equipe. Atualmente, conta com mais de 975 mil músicas, 300 mil artistas e cerca de 28 mil colaboradores. Esses dados conectam músicas baseadas em samples às suas originais, promovendo uma melhor compreensão das influências entre artistas e obras. Conforme descrito no site da plataforma, “todas as conexões musicais exibidas são contribuições de membros da comunidade WhoSampled, cuidadosamente verificadas por uma pessoa qualificada (moderadores, membros da equipe ou contribuidores certificados)”¹.

Nesta pesquisa, foram coletados dados de aproximadamente 700 mil músicas e suas conexões. A coleta foi realizada em 2019 com a devida permissão da plataforma e resultou em um grafo direcionado, onde cada nó representa uma música, e as arestas indicam suas conexões. Cada aresta é acompanhada por uma etiqueta que identifica o tipo de empréstimo musical entre as músicas conectadas. A Figura 4.2 exemplifica uma conexão do tipo *sampled* entre duas músicas, onde S representa a música que realizou o empréstimo, e T, a música original.

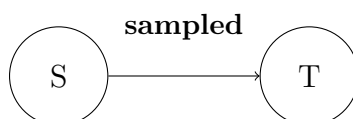


Figura 4.2: Exemplo de um empréstimo musical do tipo “sampled” entre as músicas S e T.

A Tabela 4.1 apresenta as definições dos tipos de empréstimos musicais presentes no conjunto de dados, juntamente com a frequência de ocorrência de cada um.

Cada música no conjunto de dados possui informações básicas em formato de texto, como título, ano de lançamento e artistas envolvidos, mas carece de identificadores padrão da indústria musical, como ISRC (International Standard Recording Code)², UPC (Universal Product Code)³ e EAN (European Article Number).

A ausência desses identificadores dificulta a integração dos dados do WhoSampled com outras plataformas e fontes, limitando análises que dependem de informações adicio-

¹Every musical connection we show was contributed by a member of the WhoSampled community and carefully verified by a qualified person (a moderator, member of staff or self-moderating Gold Contributor). <https://www.whoSampled.com/about>

²<https://isrc.ifpi.org>

³UPC identifica a música como um produto, seja físico ou digital

Tipo de empréstimo musical	Descrição	Ocorrências	%
sampled	S sampleou T.	291.646	32,38
is a cover of	S é uma versão cover de T.	213.317	23,68
was sampled in	S foi sampleada em T.	169.107	18,77
was covered in	S foi alvo de uma versão cover em T.	91.701	10,18
is a remix of	S é um remix de T.	85.498	9,49
was remixed in	S foi remixada em T.	49.567	5,50
Total		900.836	100,00

Tabela 4.1: Tipos de empréstimos musicais e suas ocorrências no conjunto de dados.

nais, como estatísticas de reprodução em serviços de streaming ou metadados detalhados.

4.2 Wikidata

O Wikidata é um projeto mantido pela Wikimedia Foundation, cujo propósito é oferecer uma fonte centralizada de dados estruturados e gratuitos, disponíveis para diversos fins. Trata-se de uma base de conhecimento colaborativa, organizada de forma semelhante a um banco de dados, permitindo que os usuários contribuam e acessem informações de maneira aberta e livre.

O Wikidata abrange uma ampla variedade de dados sobre pessoas, locais, eventos históricos, obras de arte e outros tópicos. Esses dados são categorizados em entidades e propriedades, proporcionando uma estrutura flexível para representar informações de forma padronizada.

A coleta de dados no Wikidata pode ser realizada de várias maneiras. A plataforma disponibiliza uma interface de consulta SPARQL, permitindo consultas personalizadas para recuperar dados específicos. Além disso, os dados podem ser acessados por meio de APIs ou exportados em conjuntos completos para uso offline. Como plataforma colaborativa, as informações no Wikidata são continuamente atualizadas e expandidas pela comunidade de editores, assegurando a disponibilidade de dados precisos e atualizados.

Para esta pesquisa, o Wikidata foi utilizado para identificar obras musicais e fornecer informações complementares como gênero musical, país de origem e códigos de identificação. Esses dados serão utilizados para referenciar de forma mais precisa as músicas na busca realizada no Google Trends.

4.3 Google Trends

O Google Trends é uma ferramenta que exhibe o interesse de busca dos usuários nos produtos do Google, permitindo a visualização dos dados em gráficos. A plataforma oferece opções de filtragem por região, produto (como imagens, notícias, compras, YouTube e pesquisa na web) e período, com dados históricos disponíveis desde 2004. Esses dados, no entanto, não são disponibilizados pela empresa por meio de APIs ou outros mecanismos que permitam consultas extensivas, sendo acessíveis apenas através do site (Figura 4.3)⁴.



Figura 4.3: Captura de Tela do Google Trends exibindo a ferramenta de análise de interesse ao longo do tempo.

De acordo com a FAQ da página oficial, “o Google Trends fornece acesso a uma amostra não filtrada de solicitações de pesquisa reais feitas ao Google. É anonimizado (ninguém é identificado pessoalmente), categorizado (determinando o tópico para uma consulta de pesquisa) e agregado (agrupado)”. Os dados fornecidos são amostras das buscas reais realizadas no Google, além de serem normalizados. A normalização facilita comparações entre termos buscados, dividindo cada ponto de dados pelo total de pesquisas realizadas em um intervalo de tempo ou localização, para representar a popularidade relativa. O resultado é escalonado em um intervalo de 0 a 100, com base na proporção de um tópico em relação a todas as pesquisas. Regiões diferentes que mostram o mesmo

⁴<https://trends.google.com.br/>

interesse de pesquisa para um termo nem sempre têm o mesmo volume total de pesquisa [Google Support Team \[2024\]](#).

A busca na plataforma pode ser feita por termos ou entidades, sendo entidades grupos de termos que compartilham o mesmo conceito, independentemente do idioma. Quando um usuário realiza uma busca em um dos produtos do Google, o sistema tenta associar essa busca a uma entidade em sua base de conhecimento (veja Figura 4.4). Essa base de conhecimento tem uma estrutura semelhante ao Freebase, adquirido pelo Google em 2010. Cada entidade na base de conhecimento possui um código de identificação exclusivo, permitindo o acesso a informações adicionais na base pública do Wikidata.

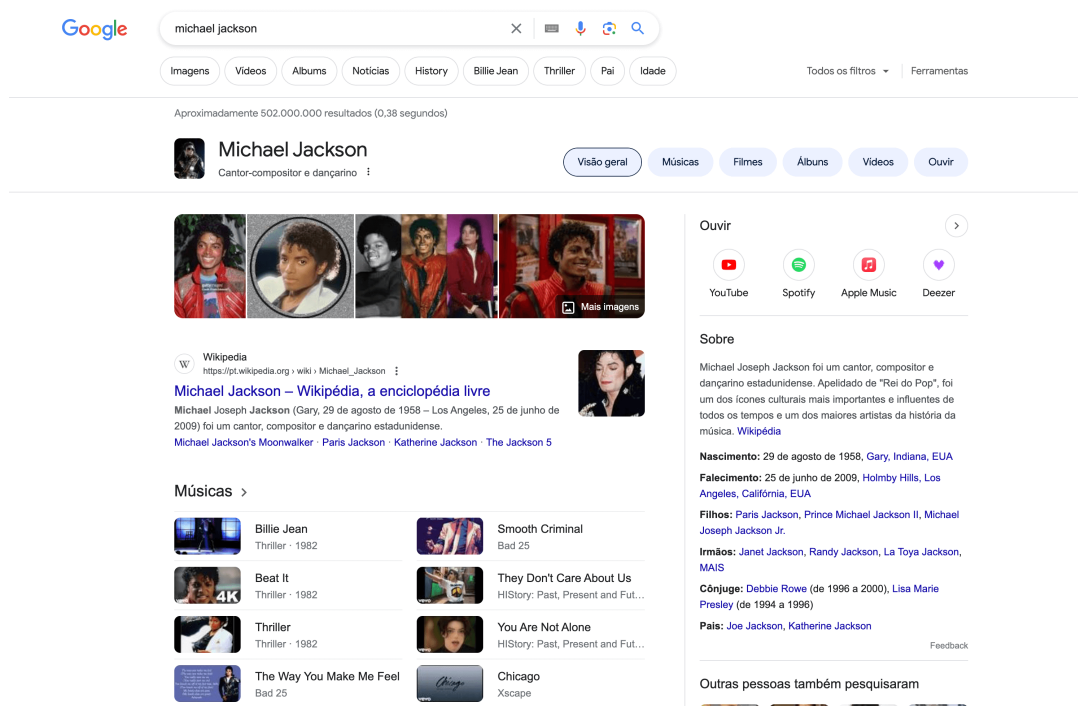


Figura 4.4: Exemplo de resultado da busca do Google exibindo informações da base de conhecimento para “Michael Jackson”, incluindo sua data de nascimento, links para plataformas de música e conexões relevantes.

Para acessar os dados do Google Trends, uma vez que não há uma API oficial disponível, faremos uso da biblioteca PyTrends⁵. Essa abordagem possui algumas limitações, como restrições de taxa de requisições e limites por endereço IP. Para contorná-las, empregamos técnicas como o uso de proxies para distribuir requisições, ajuste da taxa de requisição, estratégia de backoff exponencial e uso de IPs dinâmicos. Essas abordagens nos ajudarão a coletar os dados e contornar as restrições impostas pela plataforma.

Embora os dados do Google Trends não sejam provenientes de pesquisa científica, eles refletem o interesse de busca dos usuários em tópicos específicos e podem ser úteis

⁵<https://github.com/GeneralMills/pytrends>

quando combinados com outras fontes de dados. Nesta pesquisa, os dados serão utilizados para analisar o interesse de busca pelas músicas ao longo do tempo, permitindo mensurar o impacto do uso de samples no interesse pela música original.

4.4 Spotify

O Spotify é uma das maiores plataformas de streaming de música do mundo, com aproximadamente 551 milhões de usuários. Fundada em 2006, a plataforma oferece acesso a milhões de faixas, álbuns e playlists, permitindo aos usuários ouvirem suas músicas favoritas sob demanda, descobrirem novas músicas e criarem playlists personalizadas.

A coleta de dados do Spotify foi realizada por meio da API oficial da plataforma, que oferece diversos endpoints para recuperar informações detalhadas sobre músicas, artistas e usuários. Embora não forneça o número de execuções para cada música, a plataforma disponibiliza uma métrica de popularidade ⁶ que varia de 0 a 100, sendo 100 o mais popular. Essa métrica é calculada por um algoritmo que considera principalmente o número total de execuções e a recência dessas execuções. Assim, músicas que são tocadas com frequência atualmente têm uma popularidade maior em comparação com aquelas que foram tocadas mais no passado. A popularidade de artistas e álbuns é derivada da popularidade de suas músicas. Além disso, a API fornece informações adicionais, como a duração das músicas e o gênero dos artistas.

4.5 Youtube

O YouTube é a principal plataforma de compartilhamento de vídeos e consumo de conteúdo musical, contando com mais de 1 bilhão de usuários em todo o mundo. Esse crescimento foi impulsionado pelas iniciativas de monetização da plataforma, que incentivaram a indústria musical a disponibilizar seu material *of the Phonographic Industry [2004]*. A API do YouTube permite obter informações detalhadas sobre vídeos, incluindo dados sobre a popularidade de uma música, como número de visualizações e comentários. Essas informações são importantes para entender a audiência de uma música e podem ser usadas na análise comparativa do impacto de um sample específico em diferentes

⁶<https://developer.spotify.com/documentation/web-api/reference/get-track>

produções musicais.

Para coletar informações dos vídeos de forma rápida e escalável, utilizamos um programa de linha de comando chamado youtube-dl⁷, que nos permitiu automatizar a extração dos metadados dos vídeos no Youtube, incluindo dados de visualizações, para 699.123 músicas do conjunto de dados inicial obtido no WhoSampled. As músicas foram pesquisadas utilizando uma combinação dos campos do título da música e do nome do artista, sendo o primeiro resultado considerado como a entrada mais relevante. Em alguns casos, para contornar restrições de acesso a determinados vídeos, utilizamos cookies extraídos de uma conta Google. O Apêndice A.1 apresenta um resumo dos parâmetros extraídos para cada música, incluindo a porcentagem de preenchimento desses dados e a descrição de seu significado.

⁷<https://github.com/ytdl-org/youtube-dl>

Capítulo 5

RQ1 - Impacto em plataformas de streaming, uma visão estática

Empréstimos musicais não refletem apenas influências musicais ou sonoras, mas também carregam mensagens, cultura e estética. Cada música possui uma história única, e um empréstimo acaba incorporando parte dessa história. Neste contexto, uma questão importante no estudo da música contemporânea tem sido entender se empréstimos também transmitem parte da popularidade e do sucesso das músicas originais.

Métricas fornecidas por plataformas de streaming de música podem ser úteis ao quantificar e comparar o sucesso de músicas. Neste capítulo, buscamos responder à **Pergunta de Pesquisa RQ1**: “Como a amostragem musical impacta estaticamente a popularidade das músicas originais em plataformas de streaming, como Spotify e YouTube?”. Para isso, analisamos dados dessas plataformas, incluindo número de visualizações e métricas de popularidade, investigando a evolução do interesse por músicas originais em relação às suas versões sampleadas, covers ou remixadas, com segmentação por época e tipo de empréstimo musical.

O objetivo deste capítulo é apresentar o conjunto de dados e identificar possíveis padrões nas métricas de popularidade, comparando músicas originais e suas versões. Os resultados obtidos revelam tendências interessantes, por exemplo, versões sampleadas tendem a explorar a popularidade das músicas originais de maneira mais significativa em comparação a covers e remixes. Além disso, músicas remixadas apresentam maior variação em popularidade, sugerindo que a qualidade e a inovação no remix desempenham um papel fundamental em seu sucesso. Ao final do capítulo, discutimos as implicações desses achados para artistas e produtores musicais.

5.1 Metodologia

Para investigar o sucesso das músicas originais em comparação com suas versões, adotamos uma abordagem estruturada e sistemática. Inicialmente, descrevemos a estratégia de coleta de dados nas duas maiores plataformas de música: Spotify e YouTube. Em seguida, detalhamos o processo de organização e pré-processamento dos dados, visando garantir sua consistência e integridade. Esse processo incluiu a remoção de duplicatas, tratamento de valores ausentes e normalização dos dados, de modo a permitir comparações justas entre diferentes músicas e plataformas. Finalmente, apresentamos e discutimos os resultados obtidos por meio da análise comparativa das métricas de popularidade.

5.1.1 Coleta de Dados

Para realizar a comparação entre o sucesso das músicas originais e suas versões, enriquecemos os dados coletados no WhoSampled com informações sobre o sucesso das músicas obtidas em outras plataformas. No YouTube, obtivemos o número de visualizações dos vídeos relacionados a essas músicas e, no Spotify, uma métrica específica da plataforma que mede a popularidade das faixas. O Capítulo 4 apresenta em detalhes a estratégia de coleta adotada nestas plataformas.

5.1.2 Limpeza e preparação dos dados

Para assegurar a qualidade e precisão dos dados, realizamos duas etapas de limpeza. Na primeira, removemos os casos em que a busca textual retornou a mesma música tanto para o original quanto para a versão. Na segunda, restringimos a análise às músicas cujo título apresentava similaridade com o título registrado no WhoSampled.

Essa similaridade foi calculada com a classe `SequenceMatcher`, do módulo `difflib`¹, em Python 3. O algoritmo utilizado deriva do método de Ratcliff e Obershelp [Ratcliff et al. \[1988\]](#), conhecido como *gestalt pattern matching*, que identifica a maior subsequência

¹<https://docs.python.org/3/library/difflib.html>

contígua comum entre duas cadeias, desconsiderando elementos irrelevantes (*junk*). Esse processo é aplicado recursivamente às partes restantes, gerando índices que expressam a proximidade estrutural entre os textos. Por fim, os valores obtidos foram normalizados em escala percentual, a fim de facilitar sua interpretação.

Devido às diferenças na construção dos títulos das músicas entre as plataformas analisadas, particularmente no caso de vídeos musicais no YouTube, aplicamos limiares distintos de similaridade. A definição dos limiares foi baseada na análise exploratória da distribuição das similaridades calculadas e na avaliação qualitativa de exemplos representativos.

No caso do Spotify, observou-se que a distribuição das similaridades era altamente concentrada em torno de 100%, indicando que a maior parte dos títulos originais e das versões apresentavam alta similaridade textual. Por esse motivo, foi selecionado um limiar mais rigoroso, garantindo que apenas músicas com alta correspondência textual fossem consideradas. O limiar definido foi de 50%, de modo que 92% (293.544) das músicas coletadas permanecessem no conjunto de dados, excluindo apenas casos com discrepâncias significativas.

Por outro lado, para o YouTube, a maior variabilidade na estrutura dos títulos dos vídeos resultou em uma distribuição mais dispersa das similaridades. Essa dispersão reflete diferenças frequentes nos títulos, como a inclusão de informações adicionais (por exemplo, colaborações ou detalhes contextuais). Para lidar com essas particularidades, foi adotado um limiar mais flexível, definido como 25%. Esse valor permitiu a retenção de 91,52% das músicas (587.111), excluindo apenas entradas com baixa similaridade textual, mas preservando músicas válidas, apesar das diferenças nos títulos. A Tabela 5.1 exhibe algumas instâncias de músicas próximas ao limiar de corte incluídas no conjunto final.

Título do WhoSampled	Título do YouTube	%
God's Promises Part 1	Gospel Sermon Rev James Cleveland "GOD'S PROMISES" (Live) Nutley, New Jersey Part 1 Gospel851Music	25.21
Feds Takin' Pictures	Takin Pictures (feat. Young Jeezy, Willie the Kid, Jim Jones, Rick Ross, Young Buck & T.I.)	25.22
It's Our World	The Omen (Diamond D, Sadat X & Kamari) - It's Our World / Half Steppin' (2002 Full VLS) D.I.T.C.	25.22
Castelo Triste	Facção Central - Castelo Triste (Ft. Jota Ariais) - O Espetáculo do Circo dos Horrores - Faixa 04	25.22
Ekti Aisthisi	Xristina Koletsa - Ekti Aisthisi (Official Music Video)	25.24

Tabela 5.1: Exemplos de músicas próximas ao limiar de corte de 25% para o YouTube. A tabela apresenta títulos originais e seus correspondentes analisados, destacando casos incluídos no conjunto final, apesar de diferenças textuais significativas.

A Figura 5.1 ilustra a distribuição da similaridade de strings para cada plataforma, bem como os limiares adotados. Esses valores foram confirmados por meio de uma ve-

rificação manual em uma amostra de músicas com similaridade próxima ao limiar, assegurando que as correspondências retidas representavam adequadamente os empréstimos musicais analisados. As medidas adotadas visaram minimizar erros de identificação e garantir a integridade dos dados para as análises subsequentes.

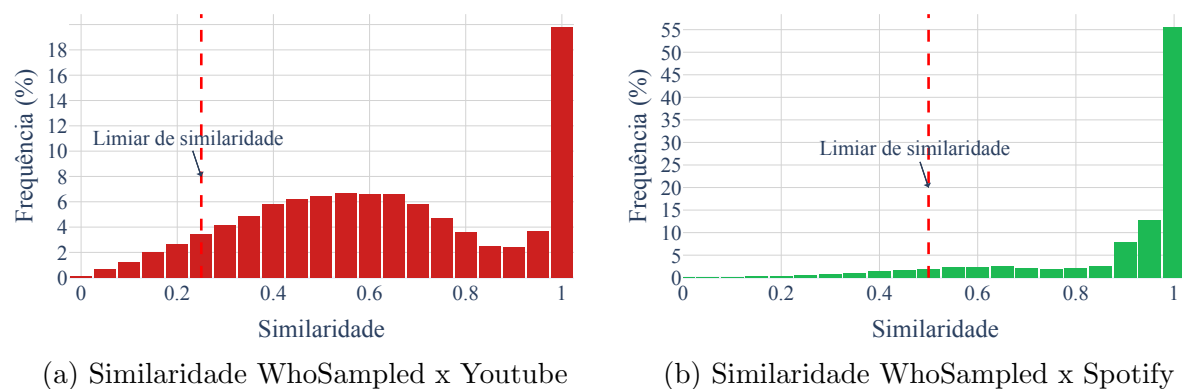


Figura 5.1: Distribuição da similaridade de strings normalizada para o nome da música

5.2 Resultados

Para responder à Pergunta de Pesquisa **RQ1**, iniciamos com uma visão geral dos dados coletados em cada plataforma e, em seguida, conduzimos uma análise comparativa das métricas de popularidade para as músicas originais e suas respectivas versões.

5.2.1 Youtube

A distribuição das músicas pelo ano de publicação (Figura 5.2) revela que a média de músicas publicadas por ano em nosso conjunto é de 35.783 (5%). Os anos com o maior número de músicas publicadas foram 2014, 2015 e 2018, respectivamente. O gráfico se comporta de forma crescente até o ano de 2011, refletindo a popularização do YouTube e sua adoção como plataforma de vídeos musicais. O decréscimo nos anos seguintes está relacionado ao intervalo de obtenção da base de dados do WhoSampled.

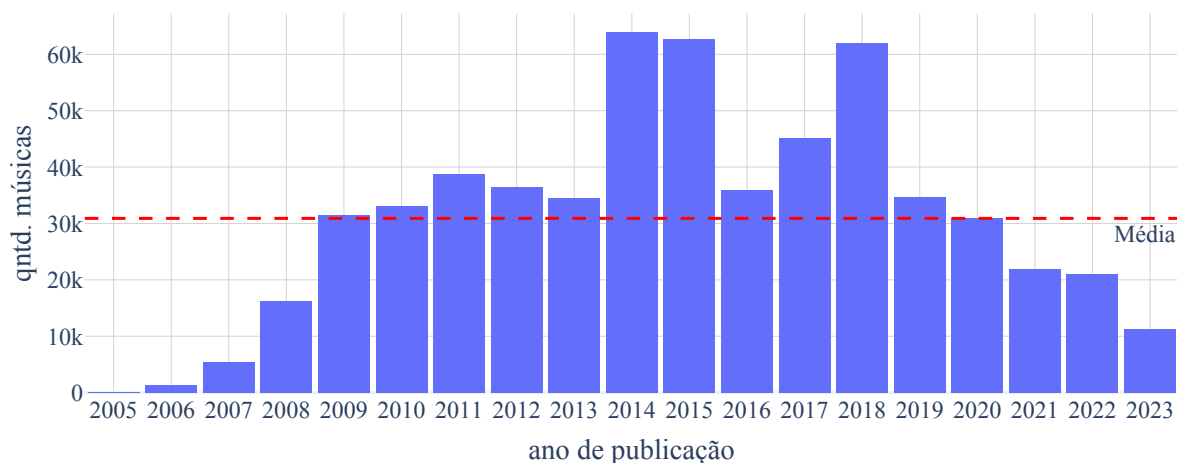


Figura 5.2: Distribuição das músicas pelo ano de publicação no YouTube.

Ao avaliarmos o número de visualizações de cada música no YouTube, observamos uma semelhança dessa distribuição com distribuições conhecidas como *power law*, caracterizadas pela predominância de uma grande quantidade de elementos com baixa ocorrência e uma pequena parcela com alta ocorrência. Esse padrão é típico de redes *scale-free*, uma categoria de redes complexas nas quais a distribuição do grau dos nós segue uma *power law*.

A Figura 5.3 mostra a distribuição do número de visualizações no conjunto de músicas. Um exemplo de concentração de visualizações é “Baby Shark Dance”, o vídeo mais popular do YouTube, com um impressionante total de 13 bilhões de visualizações. No conjunto analisado, cerca de 90% das músicas têm menos de 2 milhões de visualizações, sendo que 76% possuem até 100 mil visualizações. Entre as músicas com mais de 2 milhões de visualizações, 70% estão na faixa entre 2 e 20 milhões, e 90% têm até 100 milhões. Apenas 6% das músicas ultrapassam 200 milhões de visualizações.

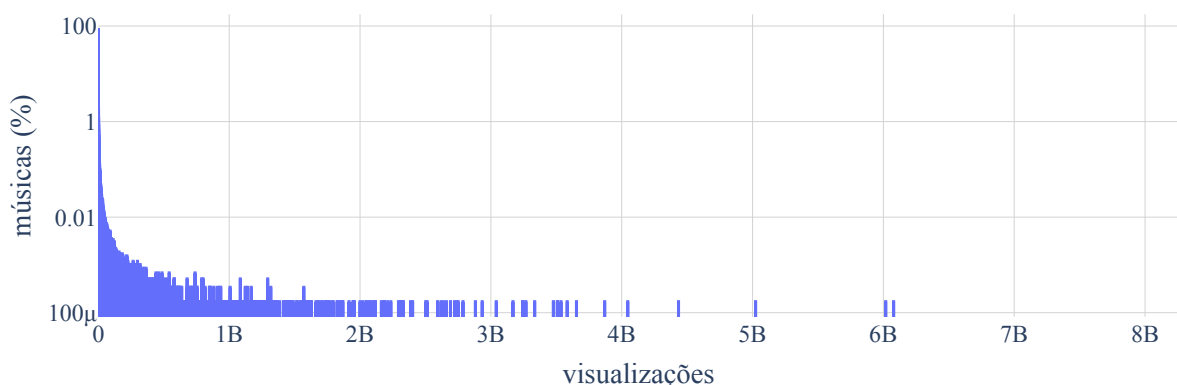


Figura 5.3: Distribuição do número de visualizações das músicas. A escala do eixo y é logarítmica e representa a porcentagem de músicas.

Para facilitar a comparação do número de visualizações entre músicas originais e suas versões derivadas (*samples*, *covers* e *remixes*), utilizamos a métrica da diferença percentual simétrica (*DPS*). Essa métrica normaliza a diferença entre as visualizações, per-

mitindo comparações que não dependem da magnitude absoluta dos valores. A fórmula utilizada é:

$$DPS = \frac{V_O - V_D}{V_O + V_D} \times 100 \quad (5.1)$$

onde V_O corresponde às visualizações da música original e V_D às visualizações da versão derivada (sample, remix ou cover).

A Figura 5.4 exhibe a distribuição dos empréstimos musicais em que a música derivada supera ou não a original em visualizações no YouTube, organizada por ano de publicação do vídeo musical na plataforma.

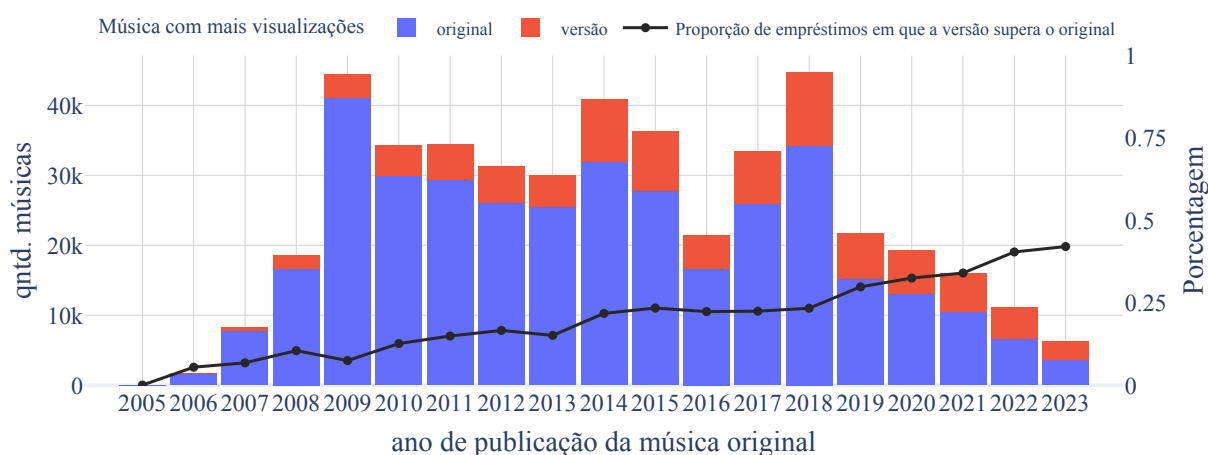


Figura 5.4: Distribuição dos empréstimos musicais em que a música derivada supera ou não a original em visualizações no YouTube, agrupados por ano de publicação.

A análise mostra que, em 19.83% dos empréstimos, a música derivada supera a original em número de visualizações ($V_O - V_D < 0$). No entanto, há um aumento contínuo na proporção de músicas derivadas que superam as originais ao longo dos anos, indicando que a prática de empréstimos musicais tem se tornado mais exitosa, no que diz respeito ao número de visualizações no Youtube.

A Figura 5.5 apresenta a Função de Distribuição Cumulativa Empírica (ECDF) das diferenças percentuais simétricas. O eixo x representa a diferença percentual simétrica, enquanto o eixo y representa a probabilidade cumulativa.

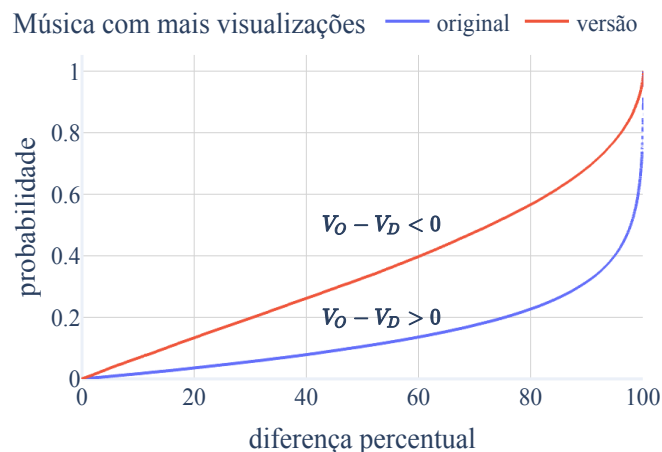


Figura 5.5: Função de Distribuição Cumulativa Empírica (ECDF) para as diferenças percentuais simétricas.

Ao analisar a ECDF, observamos duas curvas: a **azul**, que representa os casos em que a música *original* possui mais visualizações que a *derivada* ($V_O - V_D > 0$); e a **vermelha**, que representa os casos em que a *derivada* supera a *original* em visualizações ($V_O - V_D < 0$).

A ECDF mostra que, em ambas as curvas, a maioria das diferenças percentuais está próxima a 100%, sugerindo que a diferença no número de visualizações dos empréstimos musicais é, em geral, muito alta. No entanto, músicas derivadas que superam as originais (curva vermelha) possuem valores de distribuição de diferença mais uniforme, sugerindo que, nestes casos, pode haver exploração da popularidade, já que a audiência de ambas é similar.

Ao investigar empréstimos musicais com baixa diferença percentual simétrica (DPS), identificamos exemplos de exploração da popularidade. Um caso é “Das Erste Mal Im Leben” de Andreas Martin, um cover de “I Feel Like Buddy Holly” de Alvin Stardust, ambos lançados no mesmo ano. Este empréstimo pode ser ouvido na página do WhoSampled². Comparando as visualizações no YouTube, observamos uma diferença de 79.893, com o cover apresentando 4,12% menos visualizações que o original. Outro exemplo é o cover “I Love Your Smile” de DJ Antoine and Dizkodude feat. Sibbyl, com a versão original interpretada por Shanice³. Aqui, surgem limitações na metodologia, uma vez que a data de publicação no YouTube difere da data de lançamento da música original, afetando as visualizações. O original de 1991 foi postado no YouTube em 2023, enquanto o remix de 2017 já estava disponível na plataforma. A diferença é de 10.255 visualizações, representando 0,41% a mais para o cover.

A Figura 5.6 mostra a distribuição de visualizações de músicas originais e derivadas por tipo de empréstimo. Músicas originais com mais visualizações estão em azul, enquanto

²<https://www.WhoSampled.com/cover/524959>

³<https://www.WhoSampled.com/cover/519796>

as derivadas com mais visualizações aparecem em vermelho. Observa-se que a maioria das derivadas tem menos visualizações que as originais, e essa diferença é mais acentuada em *covers* e *remixes*.

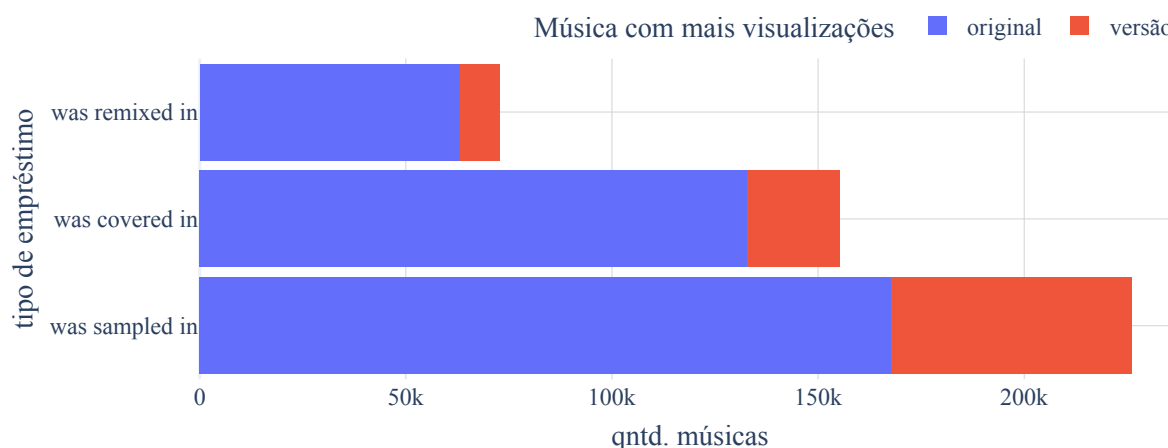


Figura 5.6: Distribuição de visualizações de músicas originais e derivadas por tipo de empréstimo musical.

Notamos anteriormente que em 19,83% dos casos a música derivada supera a original em visualizações, sendo esse percentual maior para empréstimos do tipo *sample* (Tabela 5.2). De modo geral, os resultados indicam que as músicas originais tendem a apresentar maior popularidade. É importante destacar que essas conclusões se referem ao conjunto de dados e às métricas analisadas, podendo o comportamento de consumo musical variar em outras plataformas ou contextos.

Tipo de empréstimo musical	Original com mais visualizações	Derivada com mais visualizações
cover	153,602 (85,08%)	26,931 (14,92%)
remix	72,291 (86,60%)	11,184 (13,40%)
sample	197,560 (73,08%)	72,774 (26,92%)
Total	423,453 (79,69%)	110,889 (20,31%)

Tabela 5.2: Distribuição da diferença percentual entre visualizações de músicas originais e derivadas por tipo de empréstimo.

5.2.2 Spotify

Agora, vamos aplicar a mesma análise feita no YouTube à plataforma Spotify. Embora o Spotify não forneça diretamente o número de visualizações de uma música nem a data exata de sua publicação, ele disponibiliza uma métrica de popularidade, conforme

detalhado na Seção 4.4, além da data de lançamento das faixas, que são essenciais para a nossa análise.

A Figura 5.7 apresenta a distribuição das músicas conforme o ano de lançamento registrado no Spotify, destacando uma concentração de lançamentos a partir dos anos 1990, com um pico significativo em 2019.

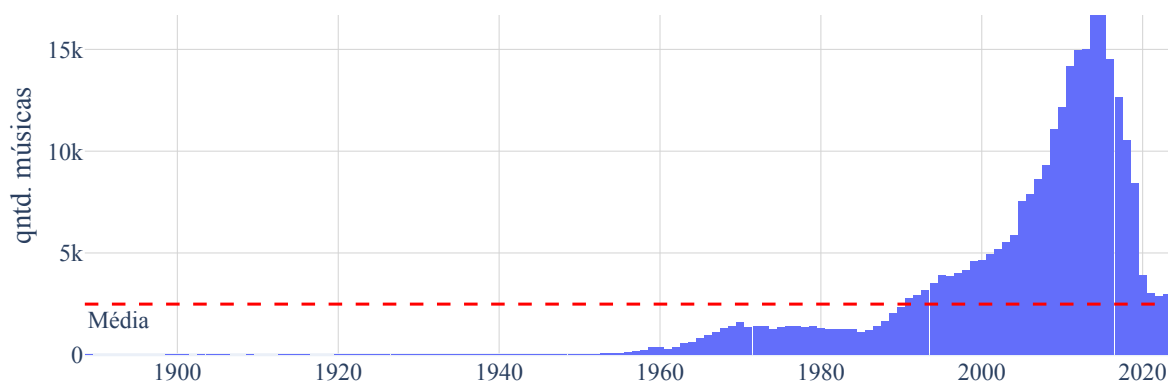


Figura 5.7: Distribuição das músicas por ano de lançamento.

Na Figura 5.8, observa-se a popularidade média das músicas ao longo dos anos de lançamento. Identifica-se uma tendência de queda na popularidade média à medida que as músicas se tornam mais recentes. Considerando que a métrica de popularidade do Spotify é ponderada pela recência das execuções, conforme discutido na Seção 4.4, os dados evidenciam que, em geral, músicas mais antigas são mais executadas. Além disso, constata-se que músicas lançadas nas décadas de 1940 e 1970 apresentaram, em média, uma popularidade superior em comparação aos outros períodos analisados.

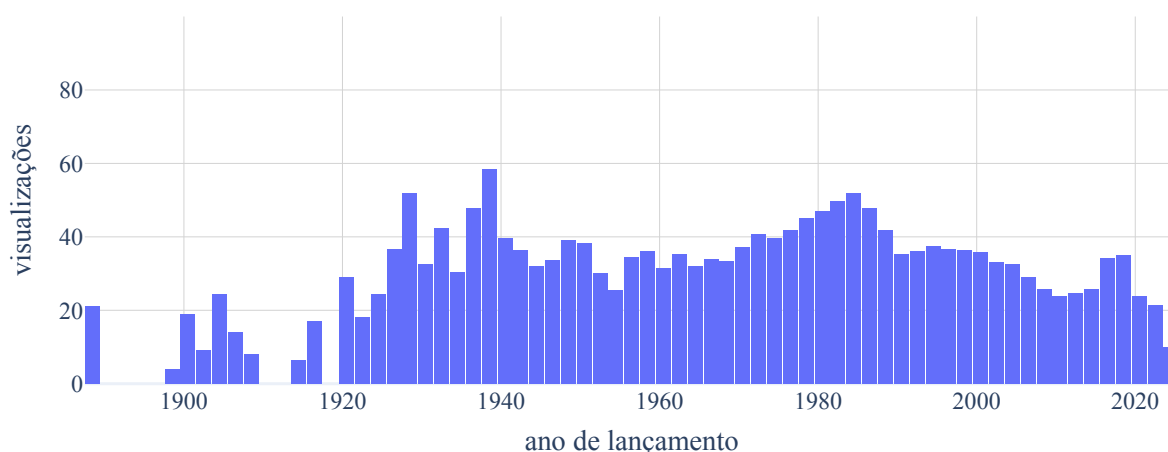


Figura 5.8: Popularidade média das músicas por ano de lançamento.

Ao analisar a distribuição da popularidade das músicas, observa-se um comportamento semelhante ao observado nos dados do YouTube, apesar das diferenças de escala, como ilustrado na Figura 5.9.

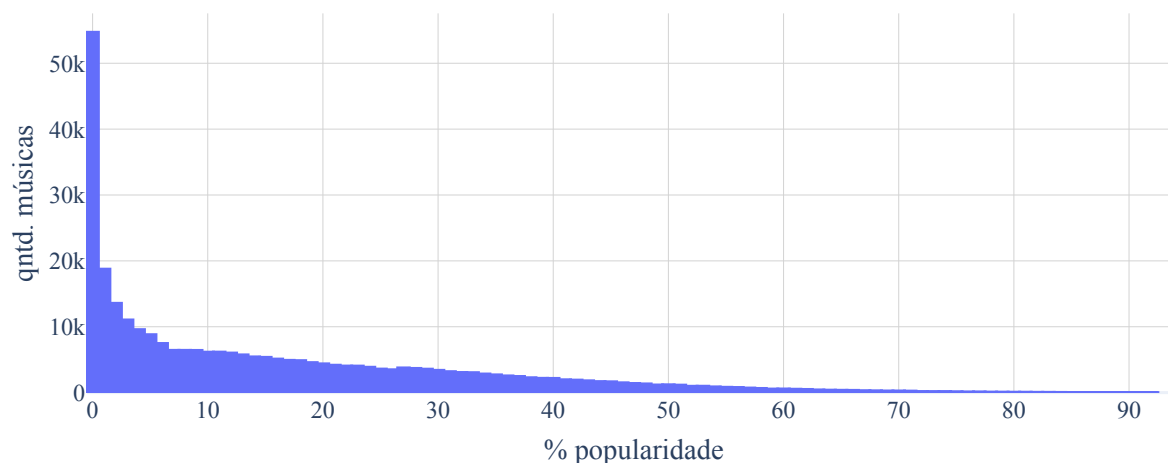


Figura 5.9: Distribuição das músicas pela métrica de popularidade do Spotify.

De maneira similar à análise do YouTube, a Figura 5.10 permite comparar a popularidade das músicas originais e suas versões ao longo do tempo, com base na métrica de popularidade do Spotify. Músicas lançadas até a década de 1950 foram ocultadas do gráfico devido ao número reduzido de amostras.

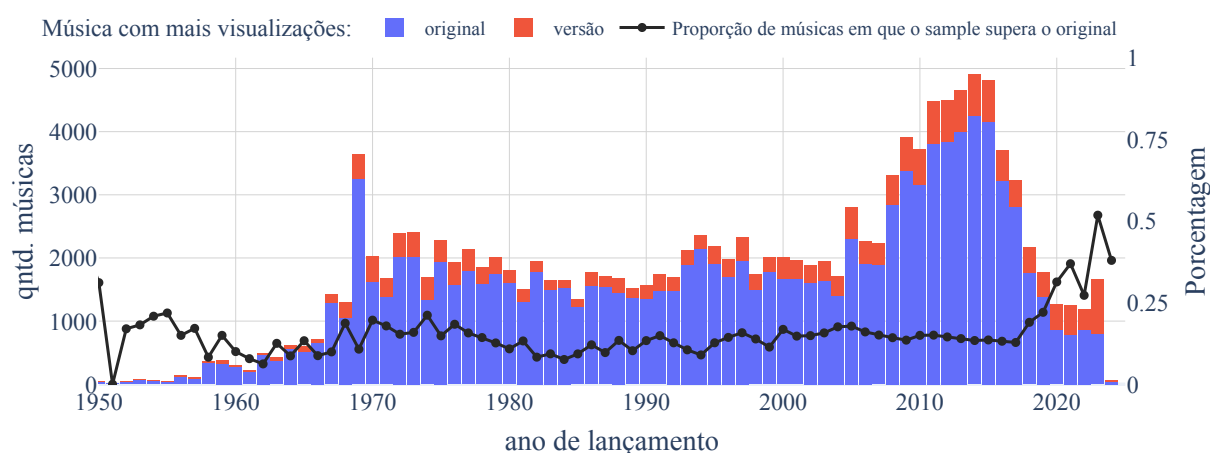


Figura 5.10: Comparação entre a popularidade da música original e suas versões ao longo do tempo, no Spotify.

Assim como no YouTube, músicas originais apresentaram maior popularidade em todas as épocas, com menos de 20% das versões superando as originais na maior parte dos anos. Essa proporção manteve-se estável ao longo do tempo, mas aumentou na última década, possivelmente em função da menor distância temporal entre lançamentos. Em 2023, pela primeira vez, versões derivadas tornaram-se mais populares do que as originais.

No Spotify, a diferença de popularidade entre originais e versões apresenta distribuição mais equilibrada, com valores positivos e negativos em proporções semelhantes. A Figura 5.11 mostra que versões mais populares tendem a superar as originais por margens reduzidas (até 40%), sugerindo que frequentemente exploram a visibilidade das músicas de origem, sem gerar grandes desvios de popularidade.

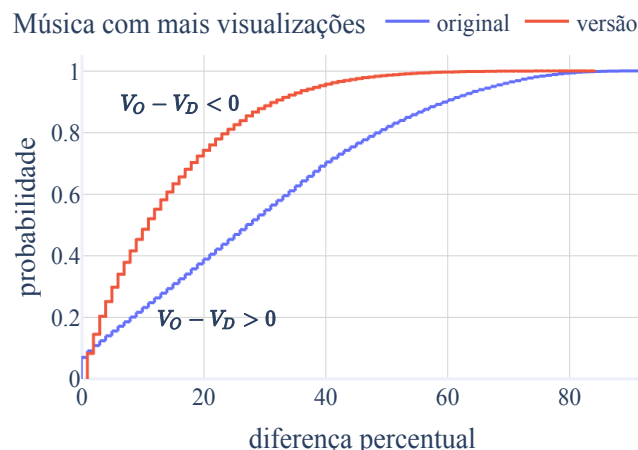


Figura 5.11: Função de Distribuição Acumulada da diferença percentual absoluta de popularidade entre a música original e a versão.

A análise comparativa segmentada por tipo de empréstimo musical mantém-se consistente, com músicas originais apresentando maior popularidade em todos os tipos de empréstimos, como ilustrado na Figura 5.12.

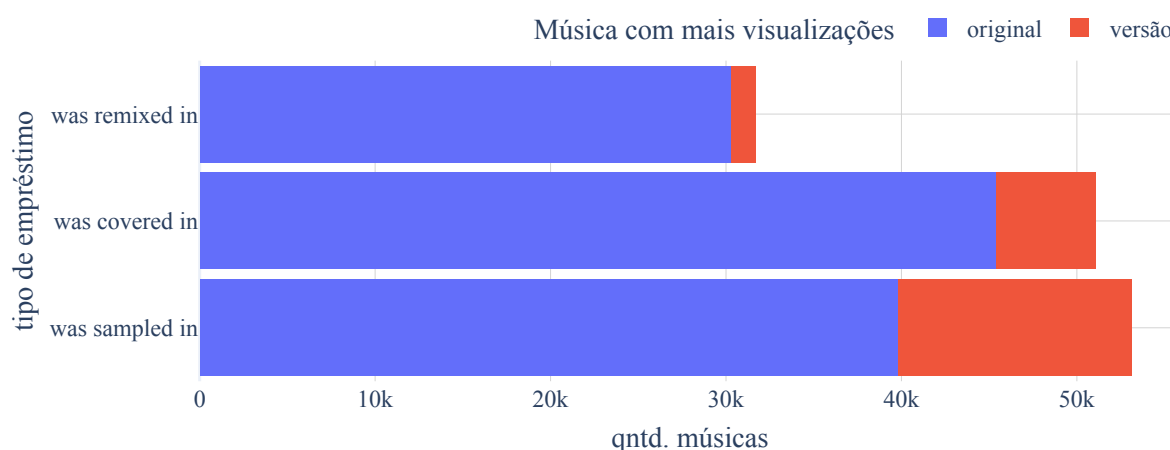


Figura 5.12: Proporção de músicas originais e versões mais populares por tipo de empréstimo musical.

A diferença no número de visualizações entre músicas originais e suas versões é mantida nas métricas do Spotify. O número reduzido de músicas pode justificar pequenas variações nos resultados observados. Mais uma vez, conclui-se que versões do tipo *sample* têm maior probabilidade de superar músicas originais em comparação com outros tipos de empréstimos musicais.

Tipo de empréstimo musical	Original com mais popularidade	Derivada com mais popularidade
cover	45,418 (88.90%)	5,671 (11.10%)
remix	30,266 (95.56%)	1,406 (4.44%)
sample	39,825 (74.96%)	13,306 (25.04%)
Total	115,509 (86.47%)	20,383 (13.53%)

Tabela 5.3: Distribuição da diferença percentual entre popularidade de músicas originais e derivadas, baseado na métrica de popularidade do Spotify.

5.3 Conclusão

Neste capítulo, apresentamos uma análise comparativa das métricas de popularidade de músicas envolvidas em empréstimos musicais. Observamos que as tendências no YouTube e no Spotify são similares, embora apresentem escalas distintas devido às características específicas de cada plataforma. Enquanto o YouTube reflete o número acumulado de visualizações de vídeos musicais até o momento da coleta, o Spotify utiliza uma métrica de popularidade percentual, ponderada pela recência das reproduções.

Os resultados mostram que, de forma geral, as músicas originais tendem a ser mais populares do que suas versões derivadas, tanto no Spotify quanto no YouTube. Essa preferência sugere que o público valoriza a originalidade ou associa um significado histórico-cultural às músicas originais, o que contribui para sua posição de destaque.

A análise da diferença percentual absoluta de popularidade indicou que, nos casos em que as músicas originais são mais populares, a diferença de popularidade é frequentemente moderada (até 50% para 80% dos empréstimos). Isso demonstra que, embora as originais mantenham vantagem, as versões derivadas também capturam uma parcela significativa de interesse do público. Além disso, a diferença de popularidade relativamente pequena entre originais e versões (raramente ultrapassando 30%) nos casos em que a versão possui mais visualizações, reforça o papel das versões como amplificadoras de visibilidade para as músicas originais, sugerindo um efeito de reciprocidade.

Ao longo do tempo, identificamos um aumento na proporção de versões que superam a popularidade das originais, embora ainda sejam casos minoritários. Esse padrão reflete possíveis mudanças na percepção do público, maior aceitação de novos formatos musicais e o papel das plataformas digitais na ampliação do alcance de versões derivadas.

Outro ponto relevante da análise é a variação de sucesso entre diferentes tipos de empréstimos musicais. Observamos que remixes tendem a se beneficiar menos da popularidade das músicas originais em comparação a covers e samples. Isso pode indicar que, para alcançar um sucesso superior, remixes exigem maior inovação ou uma abordagem criativa distinta em sua produção.

Retomando a Pergunta de Pesquisa **RQ1**, concluímos que a amostragem musical tem um impacto positivo na popularidade das músicas originais, ajudando-as a permanecer relevantes em plataformas como Spotify e YouTube. Embora as versões raramente superem as originais em popularidade, elas desempenham um papel relevante no aumento do alcance das originais. Esse efeito parece se intensificar ao longo do tempo, sugerindo um panorama dinâmico e mutuamente benéfico entre músicas originais e versões derivadas no ecossistema musical digital.

Por fim, embora essa análise forneça uma visão geral da popularidade de músicas envolvidas em empréstimos musicais ao longo de diferentes períodos, ela não mede diretamente o impacto que o lançamento de uma versão derivada tem sobre o interesse por uma obra original específica. Para preencher essa lacuna, exploramos a **RQ2**, que investiga como o empréstimo musical influencia o interesse pela música original ao longo do tempo.

Capítulo 6

RQ2 - Empréstimos Musicais e o Interesse pelo Original

Conforme discutido no Capítulo 1, alguns estudos indicam que empréstimos musicais podem impulsionar o interesse pelas músicas originais, em certos casos justificando sua classificação como uso justo. Para investigar o impacto dos empréstimos na popularidade da obra original, exploramos a hipótese de que esses elementos criam uma descontinuidade no comportamento de interesse.

Dado que a música original e o interesse do público por ela existiam antes da re-gravação, comparamos o interesse de busca pela obra antes e após o lançamento da nova versão usando Regression Discontinuity Design (RDD). Esse método permite estimar o efeito imediato da re-gravação no interesse do público pela música original. Adicionalmente, aplicamos a Causalidade de Granger para medir o impacto no longo prazo. O objetivo é identificar se o lançamento de uma nova versão resulta em aumento ou redução do interesse pela obra original, conforme definido na **Pergunta de Pesquisa RQ2**: “Como a amostragem musical influencia dinamicamente o interesse de busca pelas músicas originais ao longo do tempo?”.

Na Seção 6.1, detalhamos a coleta dos dados de interesse de busca no Google Trends. Com esses dados, realizamos uma análise de séries temporais, apresentada na Seção 6.2. Por fim, a Seção 6.3 traz a análise dos impactos de samples sobre a música original no curto e longo prazo.

6.1 Coleta de Dados

Conforme descrito no Capítulo 4, nosso estudo é fundamentado em um conjunto de dados do site WhoSampled, que mapeia relações entre aproximadamente 700 mil músicas. No entanto, esse conjunto não inclui informações de popularidade nem identificadores das músicas em outras fontes. Para obter esses dados essenciais, recorreremos a duas fontes

adicionais: Wikidata e Google Trends.

Nesta seção, detalhamos o processo de identificação das músicas nessas fontes, as limitações de nossa abordagem e o processo de limpeza dos dados.

6.1.1 Identificação das músicas no Google Trends

Para obter informações sobre o interesse de busca de cada música, realizamos dois tipos de consultas no Google Trends: por termo e por entidade. A busca por termo utiliza o nome da música e dos artistas como entrada, mas apresentou dois desafios principais. Em primeiro lugar, muitas buscas retornaram dados irrelevantes, especialmente quando o título da música inclui termos com outros significados ou há várias versões da mesma música, resultando em informações que não correspondem ao contexto musical. Em segundo lugar, várias buscas não produziram resultados, pois os usuários geralmente não utilizam muitos termos ao buscar músicas. Tentativas de refinar a pesquisa geraram consultas complexas, dificultando a obtenção de dados precisos.

Diante das limitações da busca por termos, optamos pela busca por entidades. Esta ocorre quando o usuário seleciona uma das sugestões exibidas ao digitar um termo, conforme ilustrado na Figura 6.1.



Figura 6.1: Sugestões exibidas ao digitar o termo “Clube da Esquina”. A seleção da primeira opção realiza uma busca por termo, enquanto todas as outras realizam uma busca por entidade. Fonte: Google Trends — <https://trends.google.com/trends>

De acordo com o Google, “sugestões são geradas a partir de buscas reais realizadas no Google e refletem recomendações comuns ou relevantes com base nos caracteres

digitados, além de serem influenciadas pela localização e consultas anteriores.” **Google Support Team** [2024] ¹. Ao selecionar uma sugestão, o Google utiliza um código para realizar a consulta, vinculando-se a uma entidade específica em suas bases de conhecimento. Por exemplo, ao selecionar a penúltima opção da Figura 6.1, a URL gerada é https://trends.google.com/trends/explore?q=/m/0zjw3z_, onde /m/0zjw3z_ é o ID do Freebase da música “Clube da Esquina nº 2”. A busca por entidades foi então adotada como estratégia para identificar e coletar dados históricos de interesse de busca para cada música, utilizando entidades oriundas de bases como Freebase e Google Knowledge Graph.

6.1.2 Bases de Conhecimento

Grafos ou bases de conhecimento “utilizam um modelo de dados em grafo para capturar conhecimento em cenários que envolvem a integração, gestão e extração de valor de diversas fontes de dados em larga escala” ² **Hogan et al.** [2021]. Embora o termo “Grafo de Conhecimento” tenha sido introduzido em 1972, ganhou notoriedade em 2012 com a aquisição do Freebase pelo Google. Após essa aquisição, grande parte do conteúdo do Freebase foi transferido para o Google Knowledge Graph e o Wikidata.

6.1.3 Coleta e tratamento de informações sobre músicas no Wikidata

A estrutura dos grafos de conhecimento é fundamentada em modelos de dados em grafos, onde entidades são representadas por nós e informações são registradas e interligadas por meio de arestas, que seguem o formato Sujeito-Predicado-Objeto. Esse formato associa uma propriedade a um valor específico. Por exemplo, para registrar a nacionalidade de Milton Nascimento no Wikidata, adiciona-se uma declaração à entidade **“Milton Nascimento (Q542847)”** usando a propriedade **“país de nacionalidade**

¹Autocomplete is a feature within Google Search designed to make it faster to complete searches that you’re beginning to type. The predictions come from real searches that happen on Google and show common and trending ones relevant to the characters that are entered and also related to your location and previous searches.

²Knowledge graphs use a graph-based data model to capture knowledge in application scenarios that involve integrating, managing and extracting value from diverse sources of data at large scale

(Property:P27)” com o valor **“Brasil (Q155)”**.

Para identificar as entidades relacionadas a cada música envolvida nos empréstimos musicais, realizamos uma busca textual utilizando a API do Wikidata. Seguimos os seguintes passos para cada música:

1. Realizamos três consultas de busca textual: uma apenas com o título da música, outra apenas com o nome do artista e uma terceira com a concatenação do título da música e do nome do artista. Para cada consulta, recuperamos as 10 primeiras entidades. Abaixo um exemplo da chamada realizada na API do Wikidata:

```
async function searchWikidataEntity({ track, artist }) {
  const url = 'https://www.wikidata.org/w/api.php';
  const params = {
    action: 'query', list: 'search',
    srsearch: `${track} ${artist}`, format: 'json',
  };
  return Promise.all([
    axios.get(url, { params }),
    axios.get(url, { params: { ...params, srsearch: track } }),
    axios.get(url, { params: { ...params, srsearch: artist } }),
  ]);
}
```

2. Removemos entidades duplicadas e consultamos os dados em lotes de 50 entidades. Abaixo está o método utilizado para obter as informações das entidades:

```
async function getWikidataEntity(ids) {
  return axios.get('https://www.wikidata.org/w/api.php', {
    params: {
      action: 'wbgetentities',
      ids: ids.join('|'), languages: 'en',
      format: 'json', sites: 'enwiki',
    },
  });
}
```

3. Por fim, processamos os dados coletados, extraindo as propriedades de interesse.

Para assegurar a precisão dos dados coletados em nossa pesquisa, é primordial considerar que, durante a busca textual usando o nome da música e do artista, algumas entidades retornadas podem não representar estritamente uma obra musical. Para lidar com essa questão, recorreremos à propriedade **“instance of (P31)”** no Wikidata, permitindo-nos identificar o tipo de entidade associada a cada resultado. Em particular, buscamos por valores que se enquadrem na categoria **“musical work (Q2188189)”** ou suas subclasses, como **“song (Q7366)”** ou **“composition (Q204370)”**. Essa abordagem nos possibilita filtrar as entidades que não estão diretamente relacionadas a obras musicais, garantindo assim a relevância e a precisão dos dados para nossa análise. A Tabela 6.1 descreve detalhadamente as classes utilizadas para filtrar as entidades relacionadas a obras musicais.

ID Wikidata	Nome	Descrição
Q105543609	musical work/composition	Wikidata metaclass; legal concept of uniquely identifiable piece or work of music, either vocal or instrumental; NOT applicable to recordings, broadcasts, or individual publications of music in printed or digital form or on physical media
Q2188189	musical work	generic term for any work of art related to music, i.e. songs, compositions, groups of compositions, sheet music, melodies, albums, musical films, etc.
Q55850593	music track with vocals	track on a music release that features vocals
Q207628	composed musical work	original piece or work of music, either vocal or instrumental
Q7366	song	musical composition for voice(s)
Q9734	symphony	extended musical composition
Q114798209	musical work	legal concept defining a type of copyrightable work
Q134556	single	group of single releases by an artist usually released at the same time with the same title and tracks but in different formats for consumption (digital, CD, LP)

Tabela 6.1: Classes relacionadas a obras musicais no Wikidata, utilizadas na filtragem da coleta. Fonte: <https://www.wikidata.org/>

A partir do conjunto de dados obtido no WhoSampled, contendo 699.123 músicas, coletamos um total de 1.504.008 entidades únicas. Após aplicarmos um filtro baseado nas classes listadas na Tabela 6.1, reduzimos significativamente o número de entidades para 111.238. Destas, 66.373 (aproximadamente 59,6%) não possuíam identificadores em

bases de conhecimento como Freebase e Google Knowledge Graph, resultando em um total de 44.857 músicas identificadas.

Para garantir que as entidades utilizadas nas análises representem com precisão as músicas do nosso conjunto original, aplicamos um filtro baseado na similaridade entre os nomes das músicas e artistas do conjunto de dados original e as entidades do Wikidata. Utilizamos a biblioteca `difflib`³ do Python 3 para calcular a métrica de similaridade de strings separadamente para cada nome. A Figura 6.2 ilustra a distribuição da similaridade de strings entre o conjunto original do WhoSampled, que contém as informações dos empréstimos musicais, e os dados do Wikidata, que foram utilizados na coleta de dados do Google Trends.

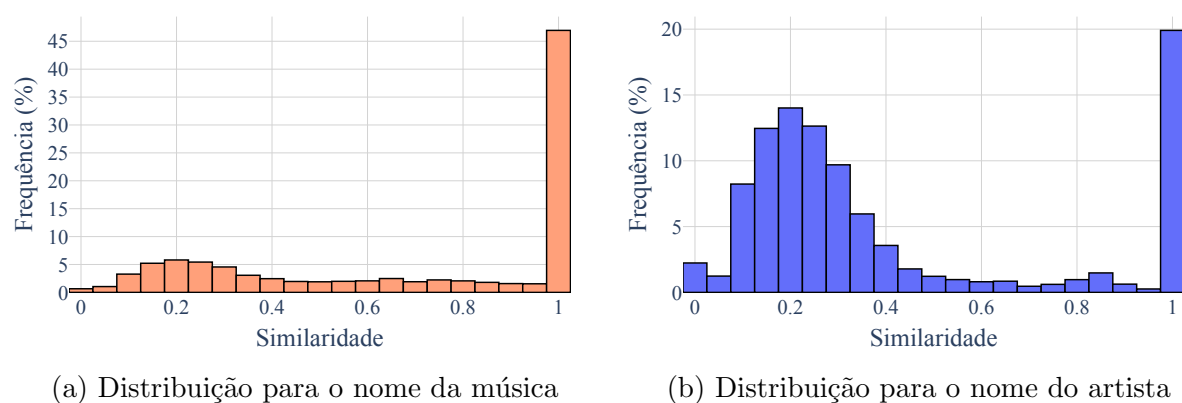


Figura 6.2: Distribuição da similaridade de strings normalizada entre informações do Wikidata e o conjunto original

Mantivemos em nosso conjunto de dados apenas músicas cuja similaridade de strings fosse superior a 55%, tanto para o nome do artista quanto para o nome da música. Esse limiar foi escolhido com base na distribuição dos valores de similaridade, especialmente para os nomes dos artistas, onde a maior parte da distribuição estava concentrada em valores abaixo desse limiar. A escolha do limiar também foi validada por uma verificação manual, utilizando uma amostra próxima ao limite, de maneira semelhante à abordagem descrita no Capítulo 3. Após a aplicação desse filtro, o número total de músicas foi reduzido em 57,5%, resultando em um conjunto final de 25.830 músicas, representando 4.477 empréstimos musicais.

³<https://docs.python.org/3/library/difflib.html>

6.1.4 Coleta do interesse ao longo do tempo no Google Trends

A coleta dos dados do Wikidata permitiu identificar as músicas nas buscas do Google Trends. Nesta seção, abordaremos a metodologia utilizada na coleta dos dados de interesse de busca ao longo do tempo no Google Trends.

Conforme apresentado no Capítulo 4, utilizamos a biblioteca PyTrends, uma API não oficial para acessar diretamente o Google Trends a partir do Python. Essa biblioteca nos permitiu fazer consultas para obter os dados de interesse ao longo do tempo para as entidades coletadas no Wikidata, conforme discutido na seção anterior. Os dados de interesse ao longo do tempo compreendem pesquisas realizadas nos produtos do Google a partir de janeiro de 2004 e são amostrados mensalmente.

Para realizar a coleta no Google Trends, desenvolvemos um script que, para cada par de músicas conectadas no grafo original (representando um empréstimo musical), realiza uma consulta informando os códigos das músicas em bases de conhecimento. Isso nos permitiu obter o interesse de busca ao longo do tempo, parametrizado de forma comparativa entre as duas entidades. A Figura 6.3 mostra um trecho do script utilizado para coleta dos dados.

```
for index, row in df.iterrows():
    fileName = f'../../data_collection/gg/{index}.csv'

    try:
        musics = [row['kg_id_from'], row['kg_id_to']]
        pytrends.build_payload(
            musics, timeframe='all', geo='', gprop='')
        interest_over_time_df = pytrends.interest_over_time()
        if not interest_over_time_df.empty:
            interest_over_time_df.to_csv(fileName)
            print(f'[Success] Writing data to {fileName}')
        else:
            print('[ERROR] No data for this term')
    except Exception as e:
        print(f'[ERROR] {e}')
```

Figura 6.3: Script de coleta das séries temporais utilizando Pytrends

Para garantir uma coleta abrangente e consistente, utilizamos os seguintes parâmetros:

- `timeframe='all'`: Indica que os dados de interesse ao longo do tempo devem ser obtidos para todo o histórico disponível no Google Trends.

- `geo=''`: Indica que a consulta não deve ser filtrada por região geográfica específica, ou seja, os dados serão globais.
- `gprop=''`: Parâmetro opcional que permite segmentar as buscas pelo tipo de pesquisa, como busca de imagens, notícias e vídeos. Neste caso, está vazio, o que significa que toda pesquisa na Web será considerada.

Durante a coleta, observamos que algumas séries temporais continham valores zerados em todos os intervalos de tempo, indicando a ausência de dados de interesse de busca para essas entidades musicais ao longo do período analisado. Essa falta de dados pode ser atribuída a várias razões, como baixa popularidade da música, falta de termos de busca associados ou problemas técnicos na coleta de dados do Google Trends. Para contornar esse problema, decidimos descartar as séries temporais que continham valores zerados em todos os intervalos de tempo.

A coleta final englobou 4.360 empréstimos musicais, sendo que cada consulta no Google Trends retornou duas séries temporais que representam o interesse de busca ao longo do tempo para as entidades musicais fornecidas como entrada. Essas séries temporais são representadas por dados que indicam a popularidade relativa das entidades em intervalos mensais. Os valores dessas séries temporais são normalizados para facilitar a comparação entre diferentes entidades e períodos de tempo.

6.2 Similaridade entre as Séries de Interesse de Busca

Nesta seção, utilizamos as métricas de similaridade descritas no Capítulo 3 para analisar as séries temporais obtidas no Google Trends, com o objetivo de avaliar se o interesse de busca pelas músicas originais ao longo do tempo está correlacionado com o interesse por suas versões derivadas, como samples, remixes ou covers. Dos 4.360 empréstimos musicais coletados, apenas 906 correspondem a músicas lançadas após 2004, sendo este o conjunto utilizado para a análise.

6.2.1 Correlação de Pearson

A correlação de Pearson entre as séries temporais foi calculada utilizando a função *pearsonr* da biblioteca *scipy* do Python, que fornece tanto o coeficiente de correlação de Pearson quanto o p-value. Correlações foram consideradas estatisticamente significativas para p-values inferiores a 0.05. Além disso, definimos um limiar de 0.7 para identificar correlações relevantes, indicando uma associação linear forte entre as séries.

Dos 906 empréstimos musicais analisados, 339 (44.7%) não apresentaram significância estatística. A Figura 6.4 exibe a distribuição dos coeficientes de correlação para os 567 empréstimos restantes. A maioria das correlações significativas (91.4%) ficou abaixo do limiar de 0.7, sugerindo baixa associação entre a popularidade das músicas originais e suas versões derivadas. Apenas 8.6% dos pares apresentaram coeficientes acima desse limiar, indicando que, nos casos em que há correlação, a relação é predominantemente positiva.

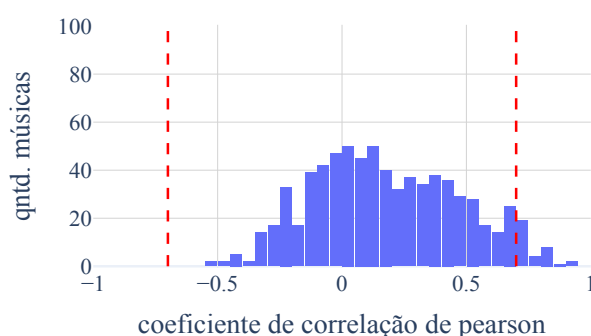


Figura 6.4: Distribuição do coeficiente de correlação de Pearson entre as séries temporais de interesse para músicas originais e suas versões.

6.2.2 Distância Euclidiana

A distância euclidiana foi usada como abordagem inicial para medir a similaridade entre as séries temporais. As distâncias variaram de 0 a 700, com 70% dos valores abaixo de 310. A Figura 6.5a ilustra a distribuição dessas distâncias. Esses resultados indicam proximidade entre muitas séries, mas a distância euclidiana apresenta limitações importantes, como a incapacidade de lidar com desalinhamentos temporais. Para superar essas limitações, adotamos o Dynamic Time Warping (DTW), apresentado na próxima

seção.

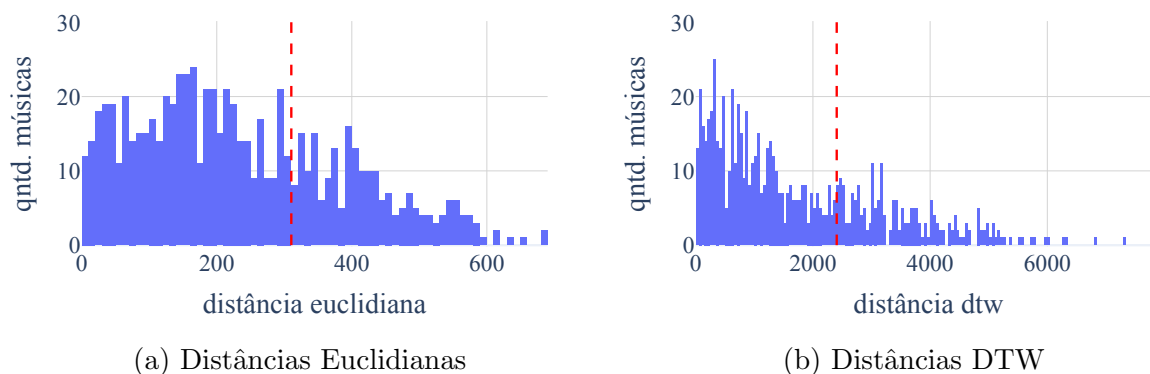


Figura 6.5: Distribuições das medidas de similaridade entre séries temporais de músicas originais e suas versões.

6.2.3 Dynamic Time Warping (DTW)

Para superar as limitações da distância euclidiana, utilizamos o método Dynamic Time Warping (DTW), implementado com a biblioteca `fastdtw` [Salvador and Chan \[2004\]](#). O DTW calcula a distância entre séries temporais considerando desalinhamentos temporais, ideal para comparar padrões deslocados no tempo.

Uma distância DTW de 0 indica que as séries são idênticas após o alinhamento. Contudo, definir um limiar absoluto para similaridade é desafiador, pois a escala depende das características das séries. Observamos que 70% das séries apresentaram distâncias DTW de até 2.396, sugerindo uma certa similaridade entre as músicas analisadas. A Figura 6.5b mostra a distribuição dessas distâncias.

Embora o DTW ofereça vantagens, como lidar com desalinhamentos temporais, ele também apresenta limitações, como a dependência de normalização e a dificuldade de interpretar suas escalas. No entanto, os resultados reforçam a hipótese de similaridade entre as séries, corroborando os objetivos deste estudo.

6.2.4 Conclusões

Os métodos aplicados indicaram uma baixa associação entre as séries temporais analisadas. A correlação de Pearson mostrou que, embora apenas 8,6% das séries apresentassem correlação significativa, a maioria delas é positiva.

As análises de distância, como Euclidiana e DTW, apresentaram distribuições concentradas na primeira metade dos gráficos, sugerindo uma certa similaridade entre as séries. Contudo, a definição de um limiar absoluto para avaliar essa proximidade continua sendo desafiadora. O DTW, especialmente, se destacou por capturar melhor as similaridades temporais, apresentando uma distribuição mais concentrada próxima de 0.

Esses resultados fornecem uma base inicial para a análise de longo prazo, abordada na Seção 6.3.2, onde será investigado o impacto das similaridades observadas por meio da Causalidade de Granger.

6.3 Impacto dos samples no interesse pela música original

Nesta seção, apresentamos os resultados da análise do impacto causado pelo lançamento de uma nova versão sobre o interesse na música original, tanto em curto quanto em longo prazo. Com uma janela temporal de 12 meses antes e depois da data de lançamento, avaliamos variações significativas no interesse de busca como indicativo de popularidade. Os resultados revelam que, em alguns casos, o lançamento de uma nova versão pode re-vigorar substancialmente o interesse pela música original, enquanto, em outros, o efeito observado é uma redução no interesse inicial.

6.3.1 Impacto no curto prazo

Utilizamos o método Regression Discontinuity Design (RDD), apresentado na Seção 3.4, para mensurar o impacto no curto prazo de um sample na música original. O RDD é um método estatístico que permite estimar o efeito causal de um tratamento em uma variável de resultado, quando a probabilidade de tratamento muda de forma

descontínua em torno de um ponto de corte. Neste caso, o tratamento é a regravação ou sample, e a variável de resultado é o interesse de busca da música original.

O RDD utiliza um ponto de corte, ou limiar, para separar as observações em dois grupos: um grupo de controle e um grupo de tratamento. Em nossa análise, o grupo de controle é composto pelo interesse de busca da música original antes do sample, enquanto o grupo de tratamento pelo interesse de busca da música original após o sample, sendo o ponto de corte a data de lançamento do sample. Para abranger um período em que fosse possível medir o efeito do lançamento do sample, definimos uma janela de tempo de 12 meses antes e depois do lançamento do mesmo para a aplicação da técnica de RDD. Em detalhe, nosso RDD foi implementado da seguinte forma:

$$googletrends(t) \sim 1 + t + \mathbb{I}_{t>0} + t \cdot \mathbb{I}_{t>0} \quad (6.1)$$

Neste contexto, t representa nossa variável de tempo, deslocada para que o tempo de lançamento da versão ocorra no tempo zero. Essa transformação é comumente utilizada em RDD, pois simplifica a interpretação dos parâmetros do modelo. O termo $\mathbb{I}_{t>0}$ é uma função indicadora que vale 1 quando $t > 0$ (após o lançamento) e zero caso contrário. Portanto, em $t = 0$, o peso deste termo reflete a diferença no intercepto, conhecida como Efeito Médio do Tratamento (ATE), uma estimativa causal do lançamento. Medimos este efeito em termos relativos (100% significa o dobro do intercepto).

Mais importante, analisamos apenas as estimativas de ATE para regressões com significância estatística em 0,05. Nosso conjunto final de dados abrangeu 617 empréstimos musicais. No entanto, apenas 82 (13%) retornaram valores p-values estatisticamente significativos. A Figura 6.6 mostra a distribuição do efeito médio do tratamento (ATE) para esses 82 empréstimos musicais com significância estatística. Discutimos esses resultados a seguir.

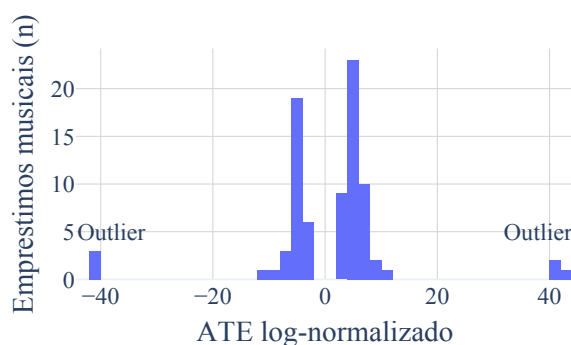


Figura 6.6: Distribuição do Efeito Médio do Tratamento (ATE) log-transformado para os 82 empréstimos musicais com significância estatística.

Inicialmente, destacamos os empréstimos com valores de ATE extremamente altos ou baixos. Em alguns casos, esses valores se devem à ausência de atividade de busca pela música original antes do lançamento da nova versão, que, ao ser lançada, despertou inte-

resse na música original. A canção com o maior ATE em nosso conjunto é *Something's Got a Hold on Me* de Etta James, de 1962, sampleada em 2011 pelo rapper Flo Rida em *Good Feeling*, como ilustrado na Figura 6.9a. Esta versão incorporou vocais e melodia de Etta James, gerando um aumento significativo no interesse pela música original, que registrou atividade de busca de 1% pela primeira vez entre 2004 e 2024. Após inspeção manual, classificamos esses eventos extremos como outliers e os desconsideramos na análise. Decidimos, portanto, focar nos casos que não são outliers, localizados no centro da distribuição (Figura 6.6).

6.3.1.1 Impacto positivo

A seguir, exploramos alguns exemplos interessantes de como a técnica RDD capturou o impacto positivo do lançamento de novas versões em alguns empréstimos musicais. A Figura 6.7 ilustra o efeito do sample da música “*Gotta Go Home*” de Boney M. (1979) na popularidade da faixa original, após ser sampleada na música “*Barbra Streisand*” pelo duo Duck Sauce (2010). A nova versão alcançou o topo das paradas de sucesso em diversos países, conquistando o Grammy de melhor música de dança em novembro de 2011. O ATE de 701,42% indica um aumento de sete vezes no interesse pela música original após o lançamento do sample, destacando como a nova versão revitalizou o interesse pelo clássico com mais de 30 anos.

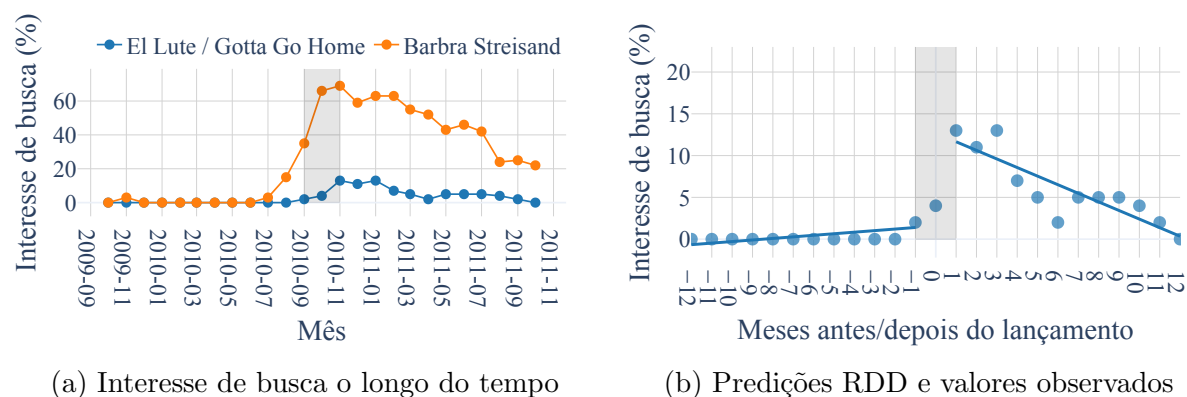


Figura 6.7: Interesse de busca por “*Gotta Go Home*” de Boney M.

Outro exemplo marcante é *Destination Calabria* de Alex Gaudino (2007), que incorpora o icônico riff de saxofone de *Calabria* de Rune RK (2003). Este riff se tornou um clássico da música eletrônica, frequentemente tocado em festivais e amplamente sampleado. Além de Gaudino, *Calabria* também foi regravada como *Calabria 2007* por Enur, com vocais de Natasja Saad, em uma versão similar à de Gaudino. O sucesso de *Destination Calabria* foi expressivo, alcançando o topo das paradas em países como Austrália e

Espanha, com certificação de platina. A análise RDD mostrou que o lançamento de *Destination Calabria* elevou o interesse por *Calabria* em 486,51% — quase cinco vezes (Figura 6.8), destacando um impacto relevante e reacendendo o interesse pela faixa original.

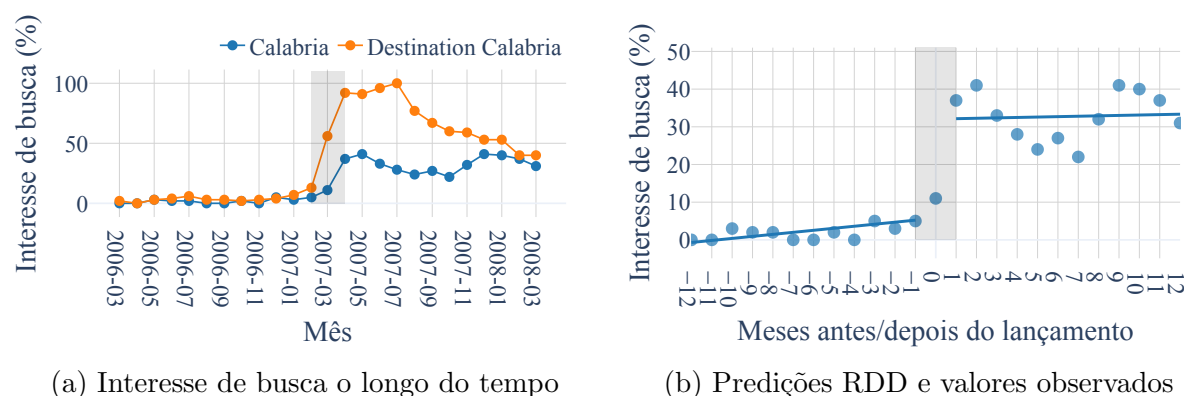


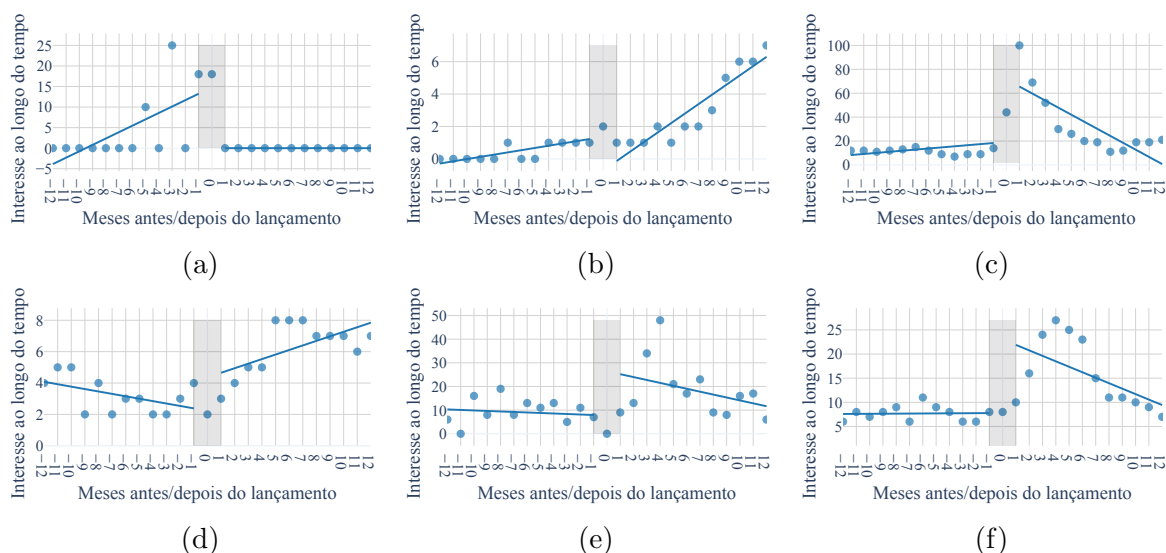
Figura 6.8: Interesse de busca por *Calabria* de Rune RK

As Figuras 6.9c, 6.9d, 6.9e e 6.9f apresentam exemplos de ATE positivo. É interessante notar que muitos desses exemplos envolvem samples que reacenderam o interesse por músicas lançadas entre o final dos anos 1970 e o final dos anos 1980, evidenciando que, em alguns casos, um empréstimo musical pode revitalizar o interesse até mesmo em obras mais antigas. No entanto, é necessário investigar se esse ressurgimento de interesse é duradouro. Nas Figuras 6.9c e 6.9f, por exemplo, observamos uma queda acentuada no interesse alguns meses após o lançamento do empréstimo. Para aprofundar essa questão, complementamos nossa análise com o estudo de Causalidade de Granger, apresentado na Seção 6.3.2.

6.3.1.2 Impacto negativo

Em cerca de 38% dos casos, o lançamento da nova versão teve um efeito negativo sobre o interesse pela faixa original. Um exemplo disso é mostrado na Figura 6.9b, que apresenta a análise RDD de *Summertime Sadness* de Lana Del Rey, sampleada em *Body Electric*, lançada cinco meses depois no EP *Paraíso*. Com um ATE de -152%, observa-se uma queda significativa no interesse pela faixa original após o lançamento do sample. Acreditamos que o marketing do EP e a similaridade entre as duas músicas tenham influenciado esse impacto negativo na popularidade da original.

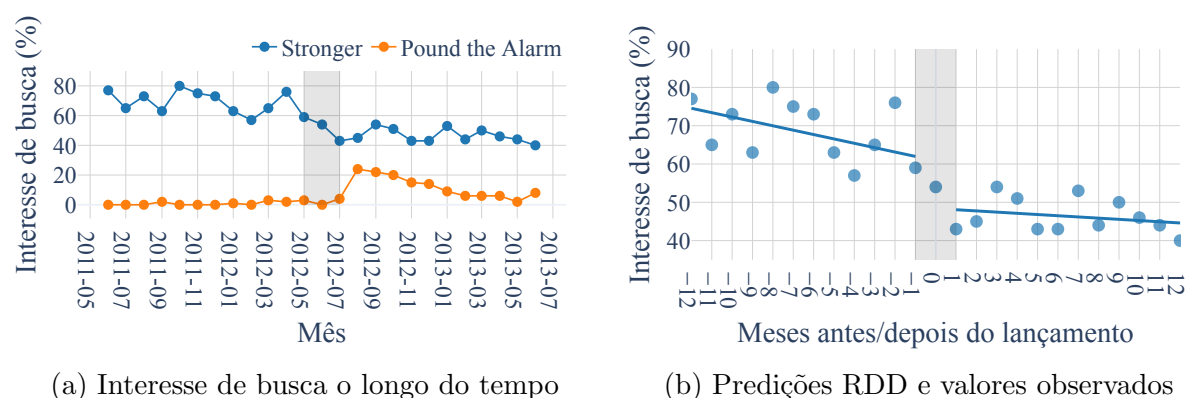
Outro exemplo de redução no interesse de busca pela faixa original ocorreu com “*Stronger*” de *Britney Spears*, lançada em 2000. Após o lançamento de “*Pound the Alarm*” de *Nicki Minaj* em 2012, o interesse de busca por “*Stronger*” começou a declinar, conforme podemos observar na Figura 6.10. Nos 12 meses anteriores ao sample, o interesse



- (a) Predições RDD e valores observados para “Something’s Got a Hold on Me” de Etta James (outlier).
 (b) “Summertime Sadness” (Lana Del Rey, 2012) emprestada por “Body Electric” (Lana Del Rey, 2012).
 (c) “You Spin Me Round (Like a Record)” (Dead or Alive, 1984) emprestada por “Right Round” (Flo’Rida, 2009).
 (d) “It Takes Two” (Rob Base DJ E-Z Rock, 1988) emprestada por “Peaches N Cream” (Snoop Dogg, 2015).
 (e) “Close to Me” (The Cure, 1985) emprestada por “So Human” (Lady Sovereign, 2009).
 (f) “Werewolves of London” (Warren Zevon, 1978) emprestada por “All Summer Long” (Kid Rock, 2007).

Figura 6.9: Exemplos de ATE

estava em aproximadamente 80%, mas diminuiu gradualmente para cerca de 55% no momento do lançamento do sample. Nos 12 meses seguintes, o interesse não ultrapassou esse patamar, eventualmente chegando ao mínimo de 40%. Esse declínio representa uma queda significativa na popularidade, com o ATE calculado em -20.47%. Além disso, a Causalidade de Granger para esse caso foi de 0.04, indicando um impacto negativo também no longo prazo, conforme veremos em mais detalhes a seguir.



(a) Interesse de busca o longo do tempo

(b) Predições RDD e valores observados

Figura 6.10: Interesse de busca por “Stronger” de Britney Spears

6.3.2 Impacto no Longo Prazo

A causalidade de Granger é uma técnica estatística que avalia se o passado de uma série temporal pode ajudar a prever o futuro de outra, levando em consideração defasagens entre observações. Diferente da correlação de Pearson, que mede apenas a relação linear entre duas variáveis, e do DTW (Dynamic Time Warping), que mede similaridade temporal, a causalidade de Granger verifica se uma série temporal fornece informações preditivas para outra, considerando os *lags* (defasagens) entre os dados.

Para realizar o teste de causalidade de Granger, utilizamos a biblioteca *statsmodels* [Seabold and Perktold \[2010\]](#), que fornece ferramentas robustas para análise de séries temporais. O teste se baseia em modelos autorregressivos (AR), comparando diferentes versões desses modelos para verificar se a inclusão de termos defasados de uma série melhora a previsão de outra. Especificamos o parâmetro *max_lag* como 10, assumindo que a relação causal entre a música original e seu empréstimo pode ser captada em até 10 meses de defasagem.

O teste de Granger retorna um *p*-value associado ao teste F de soma dos quadrados dos resíduos (SSR F-test) para cada defasagem até o *max_lag* especificado. Um *p*-value baixo (inferior a 0,05) indica que a série temporal do empréstimo contém informações preditivas sobre a série temporal da música original, sugerindo uma relação causal.

A Figura 6.11 apresenta a distribuição dos *p*-values do teste de Granger para nossa base de dados. Observamos que 64% (117) dos empréstimos musicais têm *p*-values abaixo de 0,05, mostrando que, na maioria dos casos, a nova versão (cover, sample ou remix) contém informações preditivas significativas sobre a série temporal da música original. Esse resultado sugere que os *samples* geralmente influenciam os *originais* ao longo de períodos extensos.

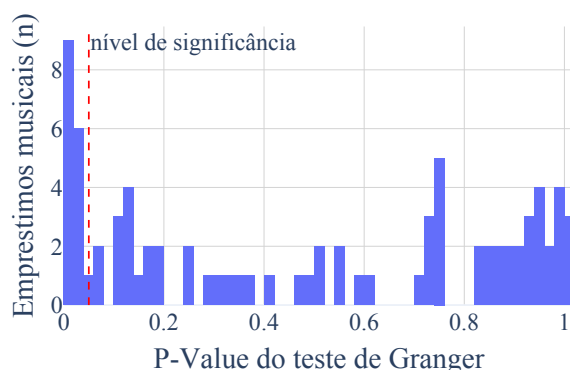


Figura 6.11: Distribuição dos *p*-valores do teste de Granger

A seguir, apresentamos exemplos de empréstimos musicais com impacto no longo prazo. No primeiro caso, “*Super High*” de Rick Ross (2010) reutiliza elementos de

“*Gangsta Gangsta*” do grupo N.W.A (1988). O teste de Granger resultou em um p -value de 0,0001, indicando que a série temporal de Rick Ross contém informações preditivas significativas sobre a série do N.W.A (Figura 6.12a).

Outro exemplo é “*All Summer Long*” de Kid Rock (2007), que reutiliza “*We-rewolves of London*” de Warren Zevon (1978). O teste de Granger também apresentou um p -value de 0,0001, destacando uma relação preditiva forte entre as séries temporais (Figura 6.12b).

Também observamos impactos duradouros entre músicas mais recentes (Figura 6.13). Dois exemplos notáveis são “*The Motto*” de Drake e “*She Will*” de Lil Wayne, ambas de 2011, e “*Rocketeer*” de Far East Movement e “*Somebody To Love*” de Justin Bieber, ambas de 2010. Em ambos os casos, o teste de Granger apresentou p -values abaixo do limiar, indicando influência preditiva significativa entre as músicas lançadas no mesmo ano.

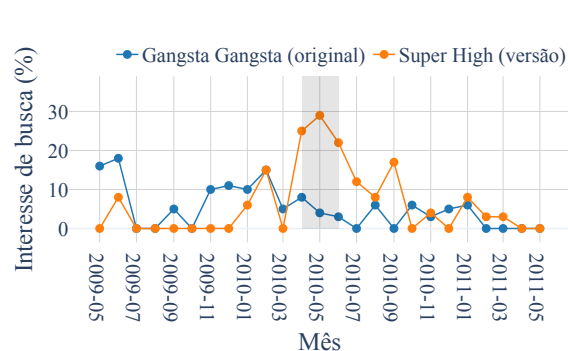
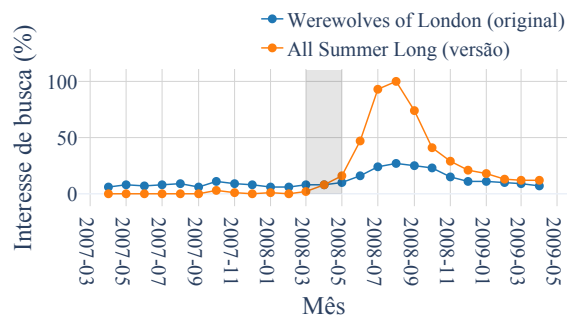
(a) “*Gangsta Gangsta*” e “*Super High*”(b) “*Werewolves of London*” e “*All Summer Long*”

Figura 6.12: Interesse de busca para músicas com alta relação causal

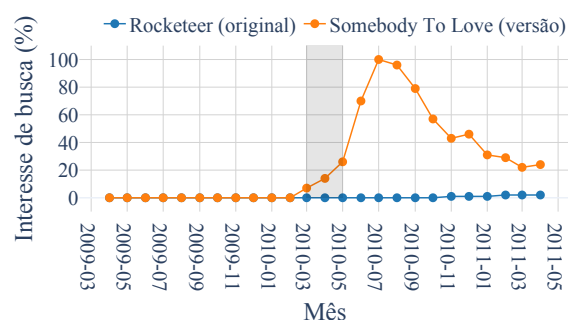
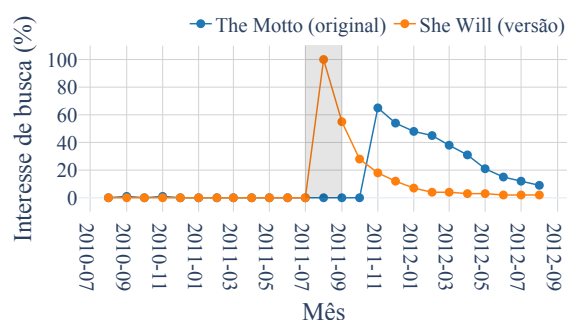
(a) “*Rocketeer*” e “*Somebody To Love*”(b) “*The Motto*” e “*She Will*”

Figura 6.13: Interesse de busca para músicas recentes com alta relação causal

Os resultados sugerem que o impacto de empréstimos musicais pode ir além do período imediatamente posterior ao lançamento, estendendo-se por meses ou até anos. Em termos interpretativos, isso pode indicar que a circulação de elementos musicais por meio de *samples*, remixes e covers está associada à manutenção ou mesmo à retomada

do interesse por obras anteriores, operando como um possível mecanismo de memória coletiva no ambiente digital. Do ponto de vista mais amplo, a persistência desse efeito também pode estar ligada a fluxos adicionais de atenção e consumo em músicas originais após seu ciclo comercial inicial, o que aponta para a possibilidade de o reuso musical criar conexões duradouras entre diferentes gerações de ouvintes.

6.4 Conclusões

Em conclusão, nossos resultados indicam que empréstimos musicais frequentemente exercem influência preditiva sobre as músicas originais, tanto em casos de composições clássicas quanto em produções mais recentes. Esses achados reforçam a hipótese de que versões derivadas podem reavivar o interesse nas músicas originais, estabelecendo uma relação duradoura entre elas.

Capítulo 7

Conclusão e Trabalhos Futuros

Este trabalho investigou o impacto de práticas como samples, covers e remixes no sucesso comercial de músicas originais, destacando a relevância desses fenômenos para a indústria musical. Essa análise é importante, especialmente em um contexto onde legislações de direitos autorais frequentemente criam barreiras para a produção baseada em samples. Apesar disso, a importância histórica dessas práticas permanece, particularmente em gêneros como hip-hop, música eletrônica e rap, nos quais elas têm um papel central na evolução cultural e artística.

Com base em dados coletados no WhoSampled, mapeamos músicas que utilizam samples, covers e remixes em suas produções. A partir desse conjunto de dados, buscamos métricas que nos permitissem avaliar o impacto dessas práticas nas músicas originais, partindo da hipótese de que a amostragem musical aumenta a popularidade das obras originais. Para isso, utilizamos duas abordagens principais: uma análise estática, que compara as métricas de popularidade no momento da coleta dos dados, e uma análise dinâmica, que avalia o impacto do lançamento de versões derivadas sobre o interesse histórico de busca pelas músicas originais.

A análise estática revelou que, embora YouTube e Spotify apresentem diferenças nas escalas e métodos de medição, as tendências observadas entre as plataformas são consistentes. De forma geral, músicas originais mantêm maior popularidade em comparação às versões derivadas, o que sugere a valorização da originalidade e um possível apelo cultural associado às obras originais. No entanto, versões derivadas também capturam uma parcela significativa do público, funcionando como catalisadoras da visibilidade das músicas originais, especialmente em casos em que a diferença de popularidade é moderada.

Observamos, ainda, um aumento na proporção de versões derivadas que superam as originais em popularidade nos últimos anos, reflexo da evolução no comportamento do público e da amplificação do alcance por meio das plataformas digitais. Entre os tipos de empréstimos musicais, remixes demonstraram maior dificuldade em alavancar popularidade quando comparados a covers e samples, sugerindo que, para obter sucesso, os remixes frequentemente demandam maior inovação e criatividade em sua produção.

Apesar de responderem parcialmente à **RQ1**, os resultados destacam a necessidade

de explorar mais detalhadamente os impactos temporais das versões derivadas, o que é abordado em **RQ2**. Essa questão examina o comportamento dinâmico do interesse por músicas originais após o lançamento de versões relacionadas.

A análise temporal do interesse de busca por músicas revelou padrões mais precisos sobre o impacto das versões derivadas no engajamento com as obras originais. Identificamos evidências de relações causais em alguns casos, com intensidades variáveis. A análise baseada em RDD sugere que o lançamento de versões derivadas pode revitalizar o interesse por faixas mais antigas, enquanto os impactos de longo prazo, avaliados pela Causalidade de Granger, indicam que empréstimos musicais podem exercer uma influência persistente na popularidade das músicas originais. Esses achados reforçam a hipótese de que samples, remixes e covers podem contribuir significativamente para o sucesso contínuo das músicas originais.

Este estudo oferece contribuições práticas, teóricas e metodológicas para o campo de pesquisa sobre música e popularidade no contexto digital. Do ponto de vista prático, a análise desenvolvida pode ser adaptada para outros tipos de séries temporais relacionadas à música. Por exemplo, pode-se investigar o impacto de métricas como o número de execuções em aplicativos de streaming, menções em redes sociais ou a exposição em rádios ao longo do tempo. Essa flexibilidade amplia o potencial de aplicação das metodologias empregadas, permitindo uma exploração mais abrangente dos fatores que influenciam a popularidade musical. No campo teórico, o estudo apresenta uma visão comparativa detalhada entre plataformas como Spotify e YouTube, elucidando as diferenças e similaridades no desempenho de músicas originais e suas versões. Essa abordagem contribui para o entendimento de como diferentes contextos e algoritmos de recomendação podem moldar o consumo musical e reforçar a relevância cultural de determinadas obras. Metodologicamente, o trabalho integra análises estatísticas robustas, como RDD e Causalidade de Granger, ao estudo de popularidade musical. Isso representa uma inovação ao aplicar técnicas comumente usadas em outras áreas para avaliar causalidade e impacto em um contexto musical. Esse avanço metodológico pode inspirar estudos futuros que busquem explorar relações causais em outros domínios da música e da cultura.

Embora este trabalho tenha oferecido uma análise sobre o impacto de empréstimos musicais no sucesso comercial de músicas originais, algumas limitações devem ser reconhecidas, particularmente no que diz respeito à coleta e à qualidade dos dados utilizados. Uma das principais limitações está relacionada à utilização do Google Trends. A plataforma impõe restrições na obtenção de dados em grande volume, o que limitou o escopo das análises realizadas. Além disso, os dados fornecidos pelo Google Trends são normalizados entre as séries comparadas, o que pode dificultar a identificação de impactos mais sutis na popularidade de músicas específicas. A identificação das entidades musicais para buscas no Google Trends também apresentou desafios. A base de conhecimento do Wikidata, utilizada para associar músicas às suas respectivas entidades, mostrou-se in-

suficiente, especialmente para músicas originadas em regiões como América do Sul, Ásia e África. Essa limitação reduziu o conjunto de dados disponível, impedindo uma análise mais detalhada sob a perspectiva regional. A utilização de uma base de conhecimento mais robusta, como a Google Knowledge Base, poderia mitigar esse problema em estudos futuros. Outro aspecto limitante foi a análise de popularidade baseada em buscas no YouTube. Embora a inclusão de vídeos não oficiais permita capturar um espectro mais amplo da popularidade musical, especialmente para artistas mais antigos, também introduz a possibilidade de contabilização de conteúdos que não refletem a produção dos artistas originais. Isso pode distorcer os resultados, sobretudo quando esses vídeos não oficiais alcançam grande audiência. Além disso, a métrica de popularidade fornecida pelo Spotify foi uma escolha necessária, devido à ausência de dados detalhados sobre o histórico ou a totalidade de execuções das músicas na plataforma. No entanto, a natureza proprietária dessa métrica, cujo método de cálculo não é divulgado, representa uma limitação metodológica que pode impactar a interpretação dos resultados. Outro ponto importante é que as métricas de popularidade consideradas abrangem apenas as três plataformas analisadas e não incluem execuções em outras plataformas de streaming, rádio ou canais alternativos. Apesar de se tratar de uma restrição, entende-se que essas três plataformas representam as maiores fontes de consumo musical atualmente, de modo que a análise ainda captura a maior parte da dinâmica de popularidade. Por fim, a construção do conjunto de dados apresentou restrições significativas. Apesar do esforço para integrar informações de fontes como Freebase e Google Trends, das cerca de 700.000 músicas analisadas, apenas 884 casos de empréstimos musicais foram identificados. Essa quantidade relativamente pequena sugere a necessidade de melhorias nos métodos de coleta e identificação de dados, o que poderia ampliar as possibilidades de análise em estudos futuros.

Em síntese, este estudo investigou o impacto de samples, covers e remixes no sucesso das músicas originais. As análises mostraram que, embora as versões derivadas não superem consistentemente as originais em popularidade, elas são importantes na amplificação da visibilidade das músicas originais. Além disso, a pesquisa revelou que esses empréstimos musicais podem revitalizar o interesse por faixas mais antigas, com efeitos persistentes ao longo do tempo.

Como forma de promover a transparência, a replicabilidade e a continuidade da pesquisa, os produtos derivados desta dissertação encontram-se disponíveis publicamente: um artigo detalhando parte das análises foi publicado no *1st Latin American Music Information Retrieval (LAMIR) Workshop* (Santos and Figueiredo [2024]), e o conjunto de dados utilizado, incluindo o grafo completo do WhoSampled, encontra-se acessível em <https://github.com/uai-ufmg/whosampleddata>, permitindo que outros pesquisadores repliquem e expandam os resultados aqui apresentados.

Referências

- Lei n.º 9 610, de 19 de fevereiro de 1998: Lei dos direitos autorais. Diário Oficial da União, February 1998. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/19610.htm. Acesso em: 30 ago. 2025.
- Olufunmilayo B Arewa. From je bach to hip hop: Musical borrowing, copyright and cultural context. *NCL Rev.*, 84:547, 2005.
- S. Bairon. *Texturas sonoras: áudio na hipermídia*. Hacker Editores, 2005. ISBN 9788586179440. URL <https://books.google.com.br/books?id=RsExvgAACAAJ>.
- Pablo Bello and David Garcia. Cultural divergence in popular music: the increasing diversity of music consumption on spotify across countries. *Humanities and Social Sciences Communications*, 8(1):1–8, 2021.
- BiObserver. Granger causality — Wikipedia, the free encyclopedia. <http://en.wikipedia.org/w/index.php?title=Granger%20causality&oldid=1257568235>, 2025. Acessado em 12-01-2025.
- Nicholas J Bryan and Gen Wang. Musical influence network analysis and rank of sample-based music. In *ISMIR*, pages 329–334, 2011.
- J Peter Burkholder. The uses of existing music: Musical borrowing as a field. *Notes*, 50(3):851–870, 1994.
- Google Support Team. Trends ajuda - perguntas frequentes acerca dos dados do google trends. <https://support.google.com/trends/answer/4365533?hl=pt-BR>, 2024. [Acessado em 03-02-2024].
- David Hesmondhalgh. Digital sampling and cultural inequality. *Social & legal studies*, 15(1):53–75, 2006.
- Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d’Amato, Gerard De Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, et al. Knowledge graphs. *ACM Computing Surveys (Csur)*, 54(4):1–37, 2021.
- Redação Hypheness. 11 músicas pop que fizeram sucesso usando samples de outras músicas. <https://www.hypheness.com.br/2020/12/>

- [11-musicas-pop-que-fizeram-sucesso-usando-samples-de-outras-musicas/](#), 12 2020. (Acessado em 28/04/2022).
- PhD Jin Cheong. Four ways to quantify synchrony between time series data. <https://towardsdatascience.com/four-ways-to-quantify-synchrony-between-time-series-data-b99136c4a9c9>, 2024. [Acessado em 01-11-2025].
- Yuchu Liu, David Issa Mattos, Jan Bosch, Helena Holmström Olsson, and Jonn Lantz. Bayesian causal inference in automotive software engineering and online evaluation, 2022. URL <https://arxiv.org/abs/2207.00222>.
- IFPI International Federation of the Phonographic Industry. Digital Music Report 2004. https://www.pro-musicabr.org.br/wp-content/uploads/2015/01/Digital_Music_Report_2014.pdf, 2004. [Acessado em 03-02-2024].
- José Luis Ortega. Cover versions as an impact indicator in popular music: A quantitative network analysis. *Plos one*, 16(4):e0250212, 2021.
- John W Ratcliff, David E Metzener, et al. Pattern matching: The gestalt approach. *Dr. Dobb's Journal*, 13(7):46, 1988.
- Stan Salvador and Philip Chan. Fastdtw: Toward accurate dynamic time warping in linear time and space. In *KDD workshop on mining temporal and sequential data*. Citeseer, 2004.
- Guilherme Soares S dos Santos and Flavio Figueiredo. Assessing the impact of sampling, remixes, and covers on original song popularity. *arXiv preprint arXiv:2411.01242*, 2024.
- Mike Schuster, David Mitchell, and Kenneth Brown. Sampling increases music sales: An empirical copyright study. *American Business Law Journal*, 56(1):177–229, 2019. doi: <https://doi.org/10.1111/ablj.12137>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/ablj.12137>.
- Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.
- Nuttiiya Seekhao. How streaming services changed collaboration and remix trends. <https://blog.chartmetric.com/the-evolving-role-of-music-artist-collaborations/>, 1 2020. (Acessado em 28/04/2022).
- Uri Shalit, Daphna Weinshall, and Gal Chechik. Modeling musical influence with topic models. In Sanjoy Dasgupta and David McAllester, editors, *Proceedings of the 30th*

International Conference on Machine Learning, volume 28 of *Proceedings of Machine Learning Research*, pages 244–252, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR. URL <https://proceedings.mlr.press/v28/shalit13.html>.

Romain Tavenard. An introduction to dynamic time warping. <https://rtavenar.github.io/blog/dtw.html>, 2021.

William W. S. Wei. *Time Series Analysis: Univariate and Multivariate Methods*, volume 2. Pearson Education, Inc, 2006. ISBN 0-321-32216-9.

Apêndice A

Resumo dos dados obtidos no Youtube

Parâmetro	Tipo	Descrição	%
content_id	string	Identificador único do conteúdo	100
id	string	ID do vídeo no YouTube	100
title	string	Título do vídeo no YouTube	100
duration	int	Duração do vídeo em segundos	100
view_count	int	Número de visualizações do vídeo	100
tags	string	Tags associadas ao vídeo	100
categories	string	Categorias do vídeo	100
upload_date	date	Data de upload no YouTube	100
alt_title	string	Título alternativo do vídeo	31
track	string	Nome da música	31
artist	string	Nome do artista	31
album	string	Nome do álbum	31
release_year	year	Ano de lançamento da música	31
release_date	date	Data de lançamento da música	28

Tabela A.1: Informações disponíveis nos dados provenientes do Youtube

Apêndice B

Interesse de Busca vs. RDD Predito

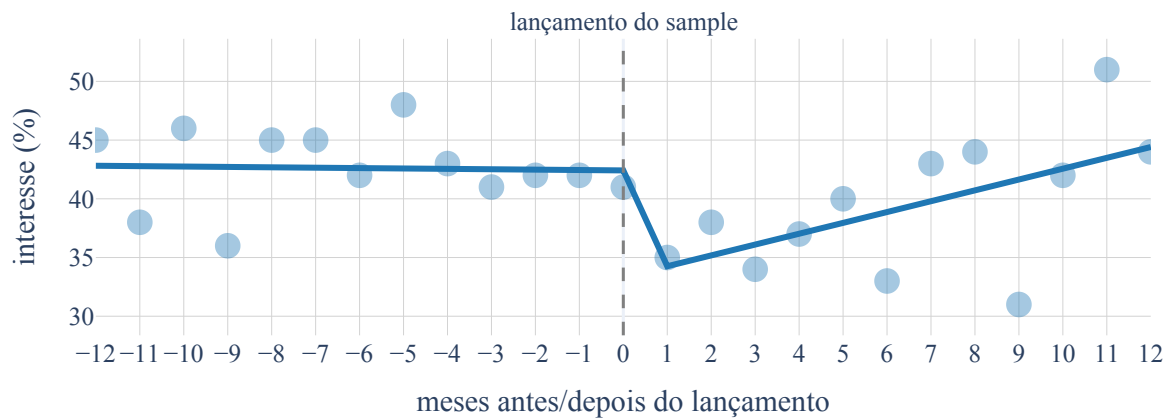


Figura B.1: Kryptonite (3 Doors Down, 1999) sampleado em Superhuman (Chris Webby, 2014). ATE Normalizado: -3.11 — Causalidade de Granger: 0.32

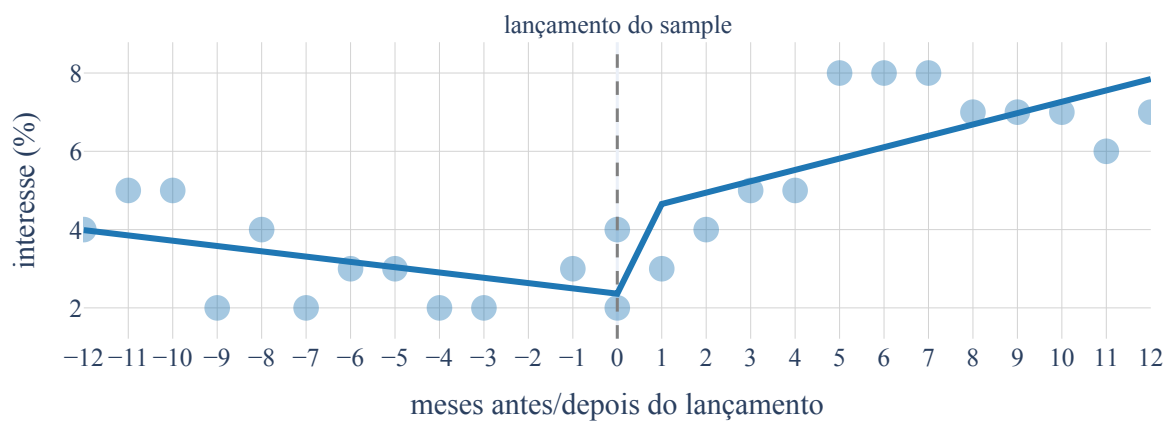


Figura B.2: It Takes Two (Rob Base & DJ E-Z Rock, 1988) sampleado em Peaches N Cream (Snoop Dogg song) (Snoop Dogg, 2015). ATE Normalizado: 4.451 — Causalidade de Granger: 0.18

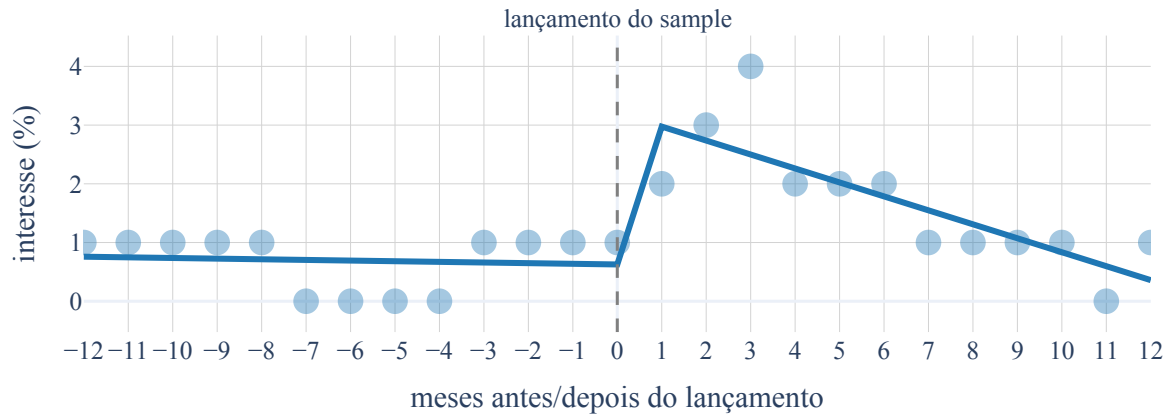


Figura B.3: U Can't Touch This (MC Hammer, 1990) sampleado em Dance (Ass) (Big Sean, 2011). ATE Normalizado: 6.025 — Causalidade de Granger: 0.84

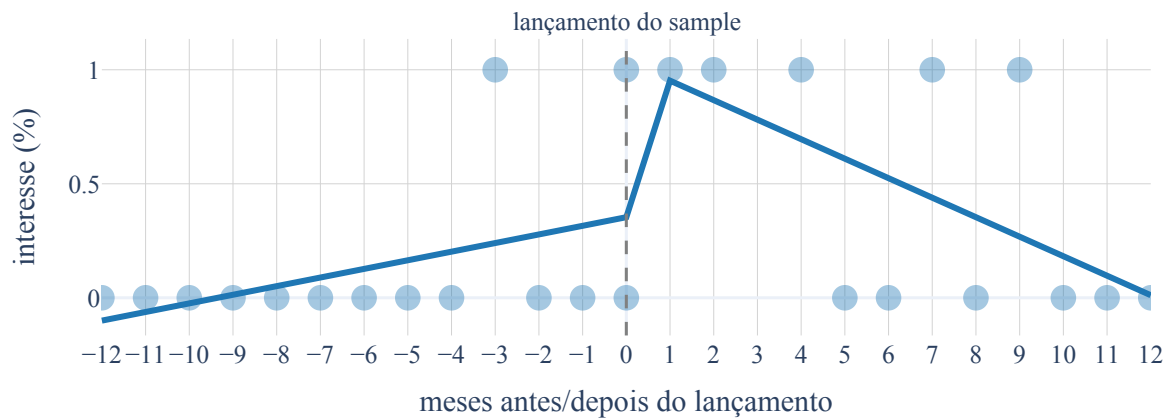


Figura B.4: Can You Feel It (The Jacksons, 1980) sampleado em Sorry (Madonna, 2005). ATE Normalizado: 5.271 — Causalidade de Granger: 0.75

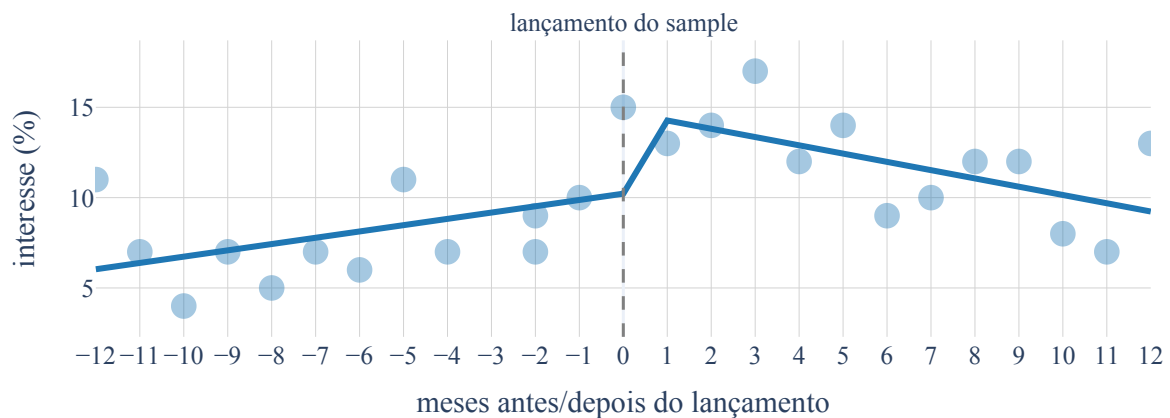


Figura B.5: Get Up Offa That Thing (James Brown, 1976) sampleado em Luv (Apink, 2014). ATE Normalizado: 3.809 — Causalidade de Granger: 0.84

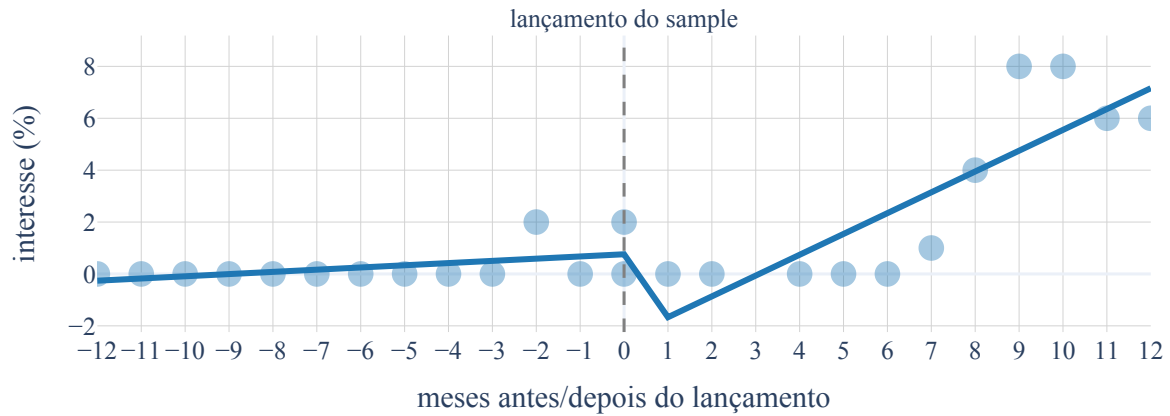


Figura B.6: Bon, Bon (Pitbull, 2010) sampleado em We No Speak Americano (Yolanda Be Cool, 2010). ATE Normalizado: -6.056 — Causalidade de Granger: 0.27

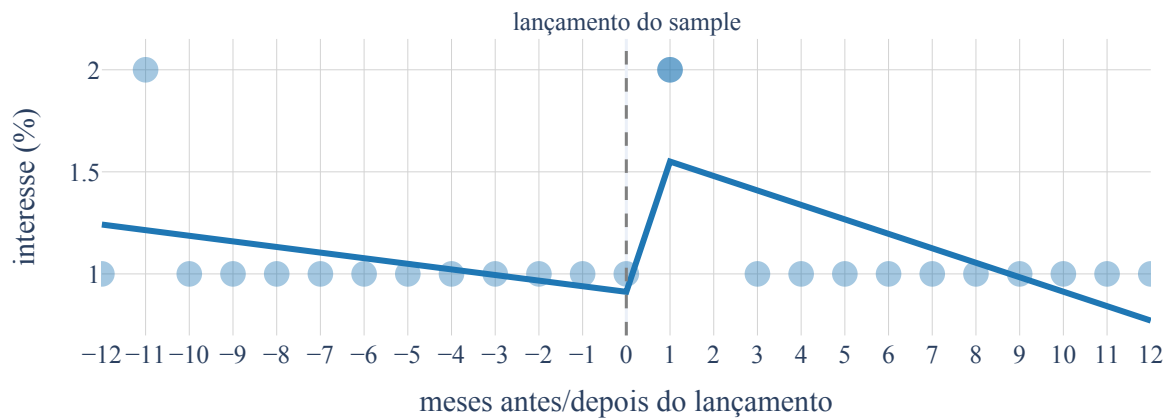


Figura B.7: Real Love (Mary J. Blige, 1992) sampleado em Super Rich Kids (Frank Ocean, 2012). ATE Normalizado: 4.366 — Causalidade de Granger: 0.99

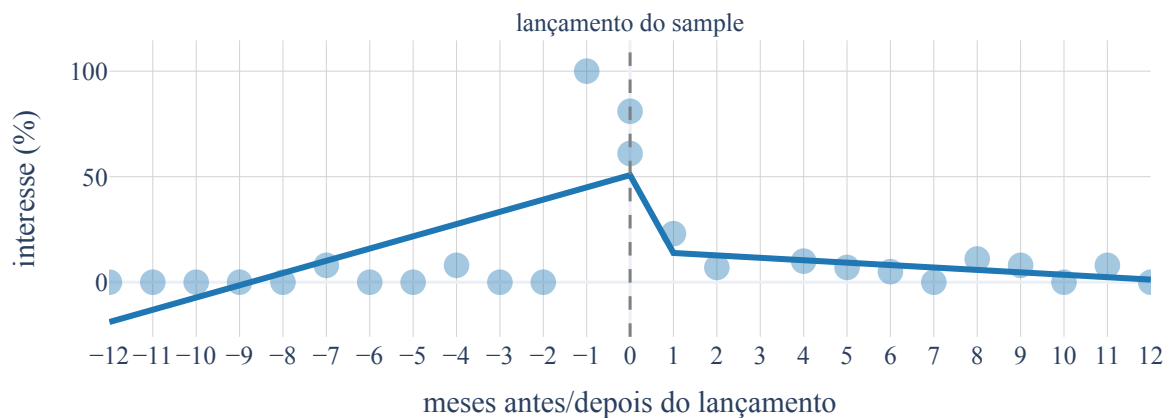


Figura B.8: Get On Your Boots (U2, 2009) sampleado em Fez - Being Born (U2, 2009). ATE Normalizado: -4.269 — Causalidade de Granger: 0.85

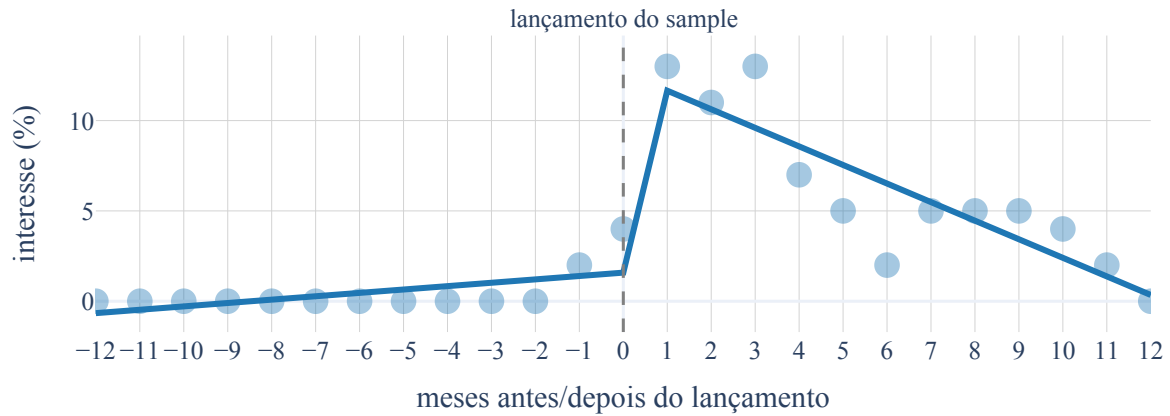


Figura B.9: El Lute / Gotta Go Home (Boney M., 1979) sampleado em Barbra Streisand (Duck Sauce, 2010). ATE Normalizado: 6.555 — Causalidade de Granger: 0.12

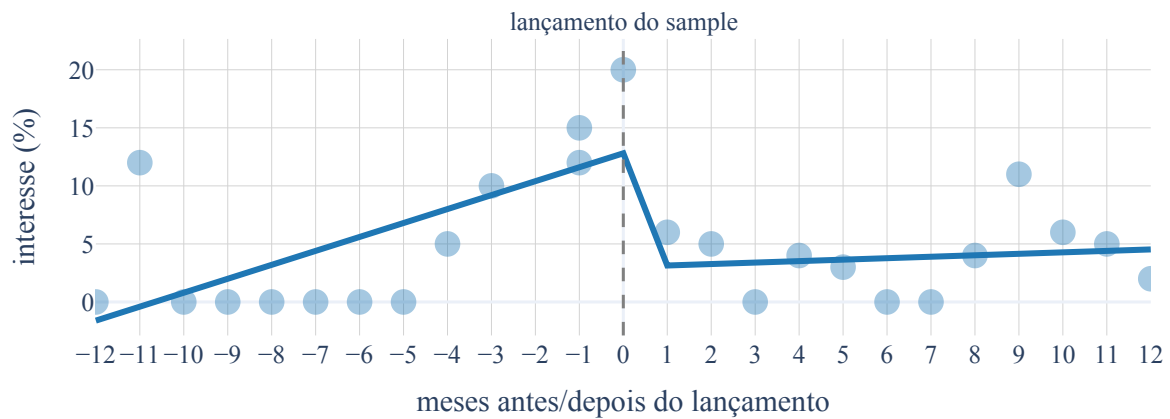


Figura B.10: Hush, Little Baby (Traditional Folk, 1939) sampleado em Mockingbird (Eminem, 2004). ATE Normalizado: -4.35 — Causalidade de Granger: 0.83

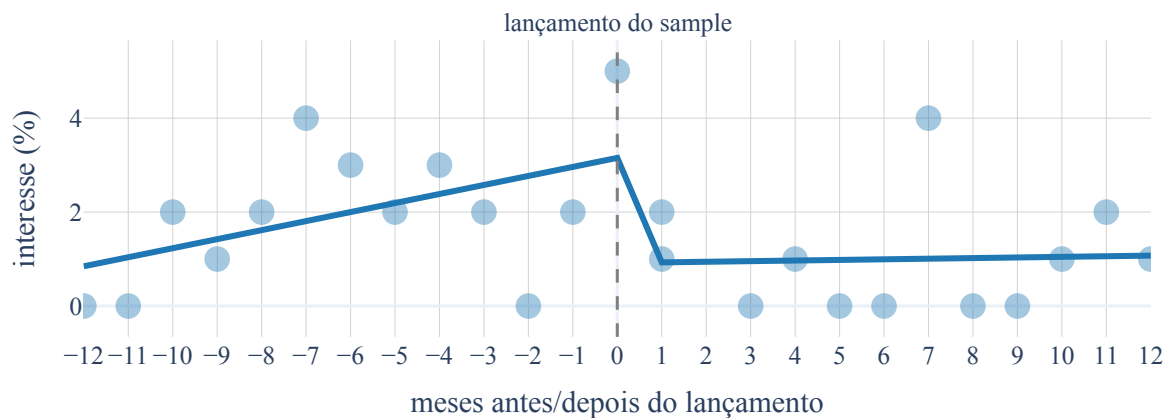


Figura B.11: Mo Money Mo Problems (The Notorious B.I.G., 1997) sampleado em Worst Behavior (Drake, 2013). ATE Normalizado: -4.276 — Causalidade de Granger: 0.74

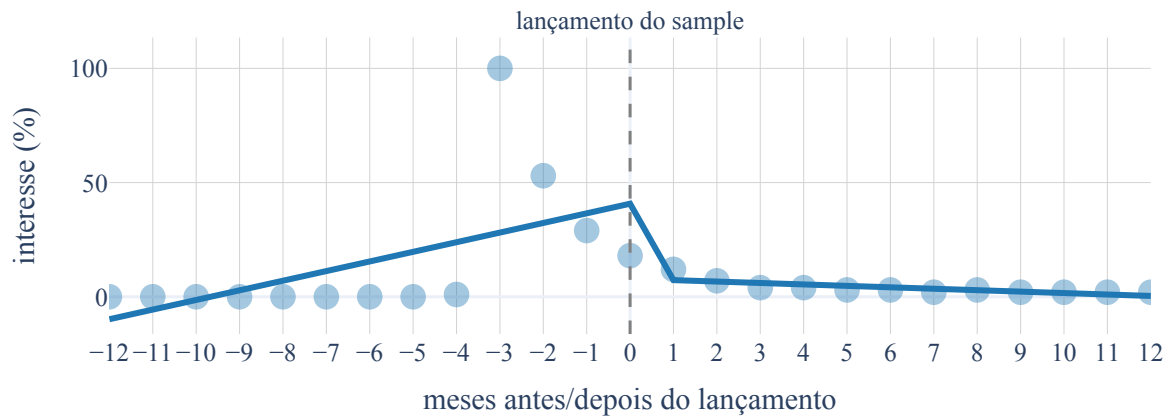


Figura B.12: She Will (Lil Wayne, 2011) sampleado em The Motto (Drake, 2011). ATE Normalizado: -4.401 — Causalidade de Granger: 0.02

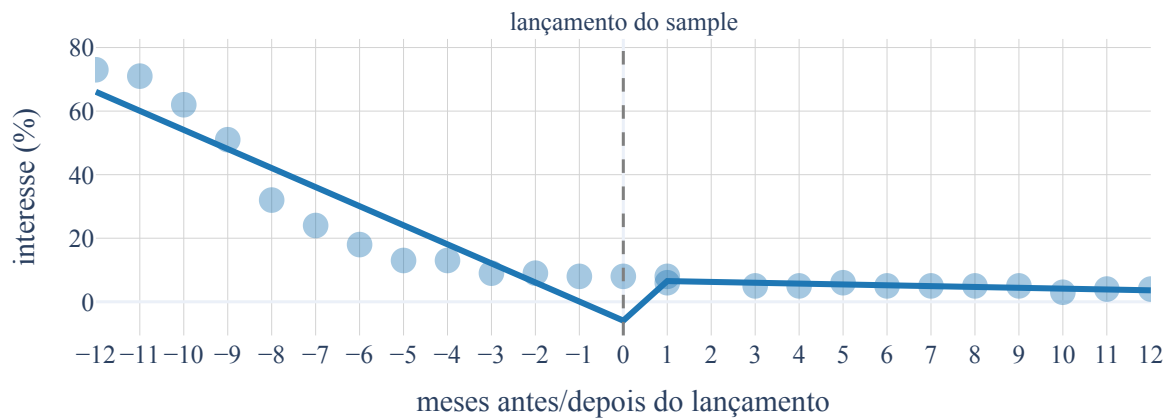


Figura B.13: The Motto (Drake, 2011) sampleado em YOLO (The Lonely Island, 2013). ATE Normalizado: -5.373 — Causalidade de Granger: 1.0

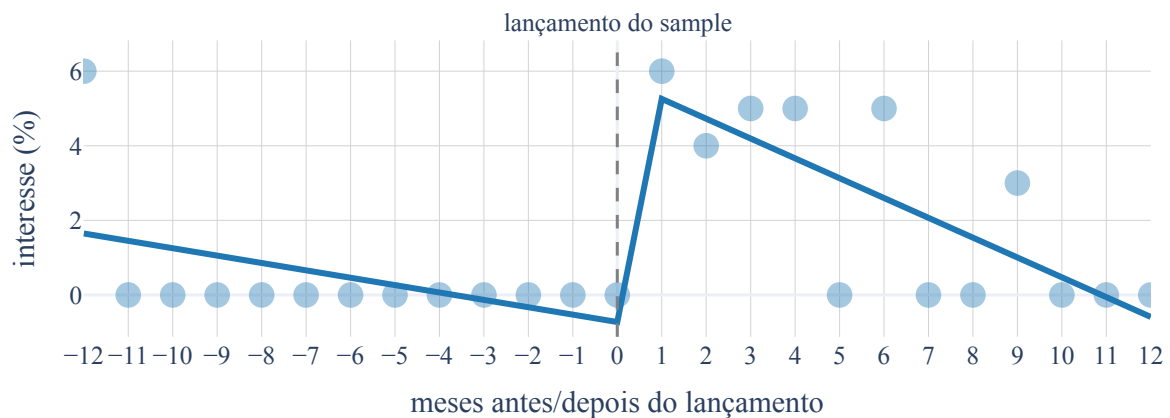


Figura B.14: Chaconne in F minor (Johann Pachelbel, 1680) sampleado em What If (Jason Derülo, 2010). ATE Normalizado: -6.801 — Causalidade de Granger: 0.11

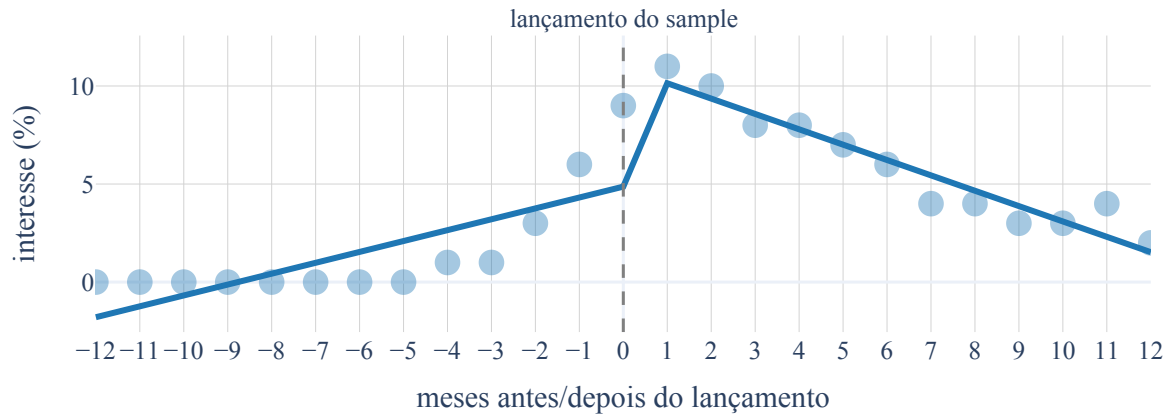


Figura B.15: Stereo Hearts (Gym Class Heroes, 2011) sampleado em IDGAFOS (Dillon Francis, 2011). ATE Normalizado: 4.832 — Causalidade de Granger: 0.02

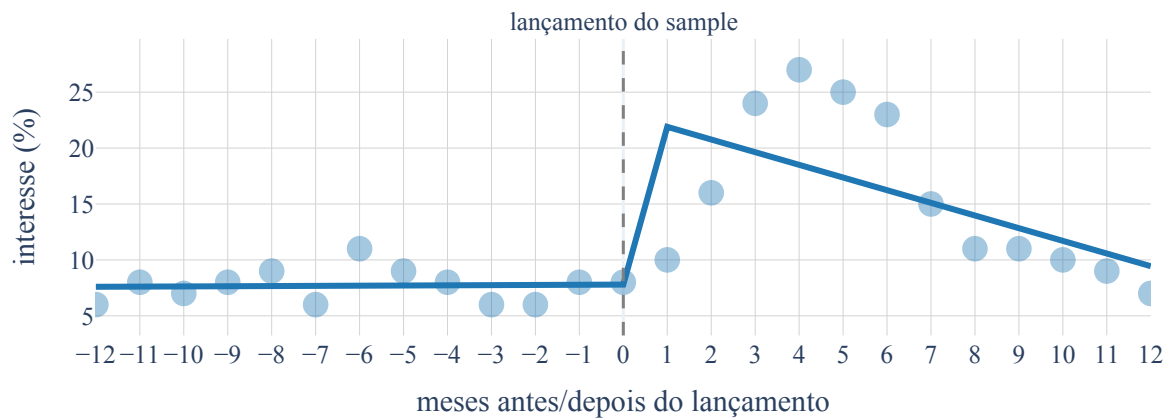


Figura B.16: Werewolves of London (Warren Zevon, 1978) sampleado em All Summer Long (Kid Rock, 2007). ATE Normalizado: 5.281 — Causalidade de Granger: 0.0

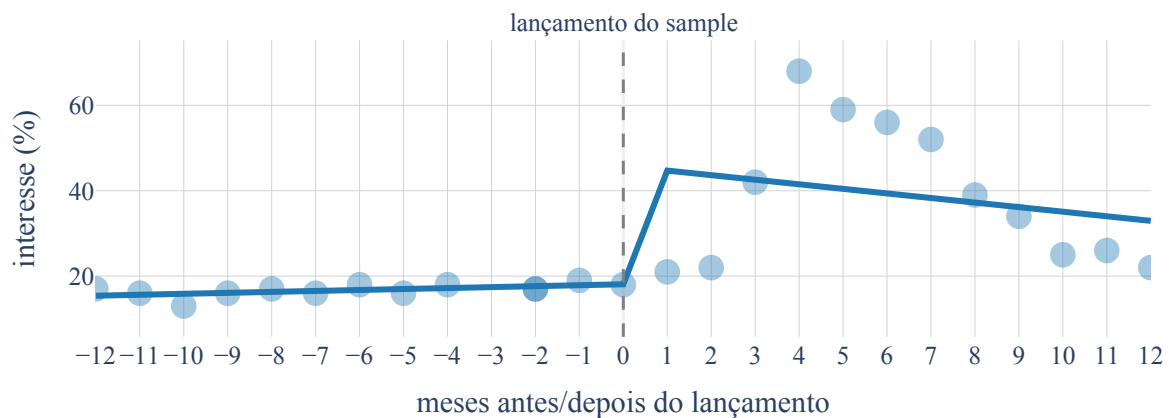


Figura B.17: Hide and Seek (Imogen Heap, 2005) sampleado em Whatcha Say (Jason Derülo, 2009). ATE Normalizado: 5.037 — Causalidade de Granger: 0.96

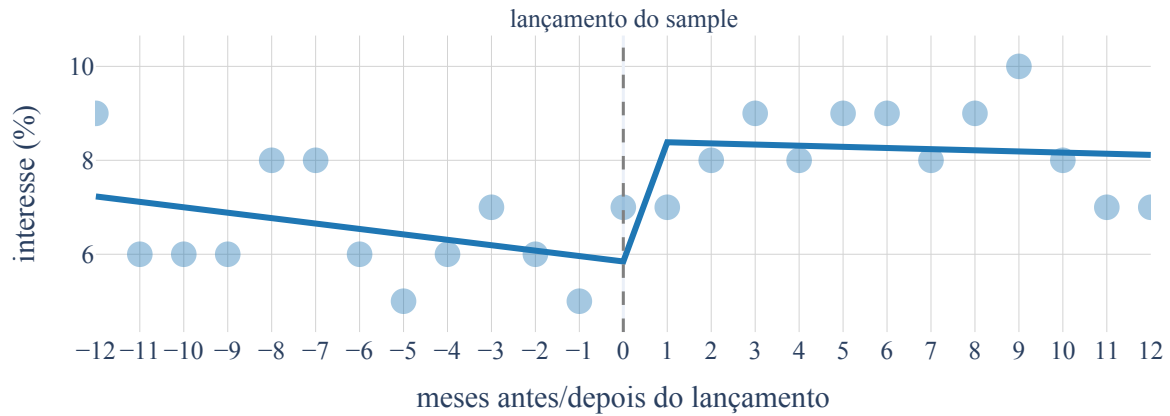


Figura B.18: Just a Friend (Biz Markie, 1989) sampleado em Say You're Just a Friend (Austin Mahone, 2012). ATE Normalizado: 3.803 — Causalidade de Granger: 0.0

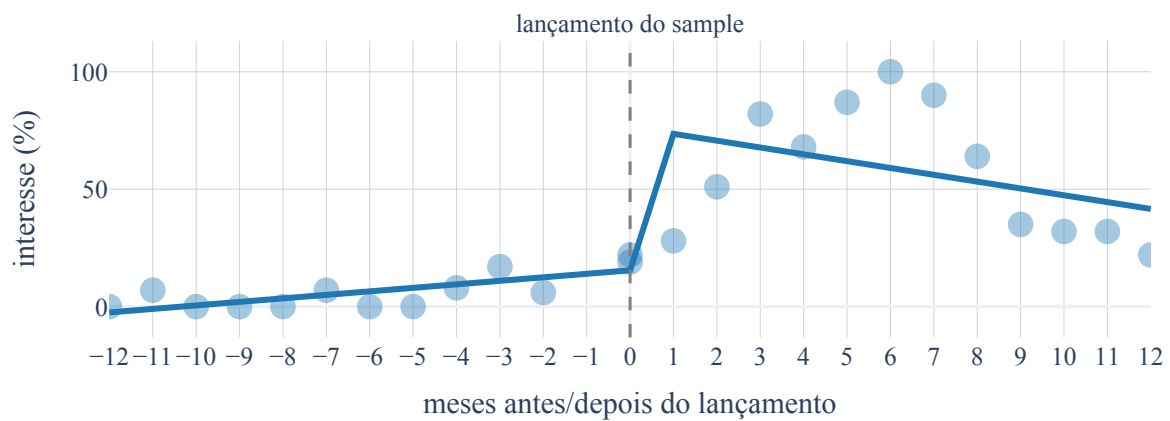


Figura B.19: One Step at a Time (Jordin Sparks, 2007) sampleado em All I See (Kylie Minogue, 2007). ATE Normalizado: 5.98 — Causalidade de Granger: 0.11

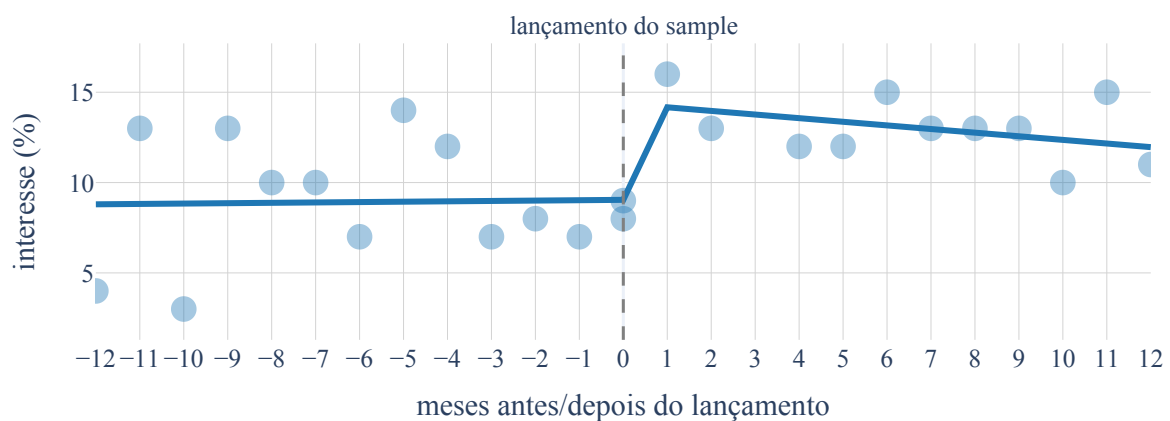


Figura B.20: Close to Me (The Cure, 1985) sampleado em So Human (Lady Sovereign, 2009). ATE Normalizado: 4.092 — Causalidade de Granger: 0.07

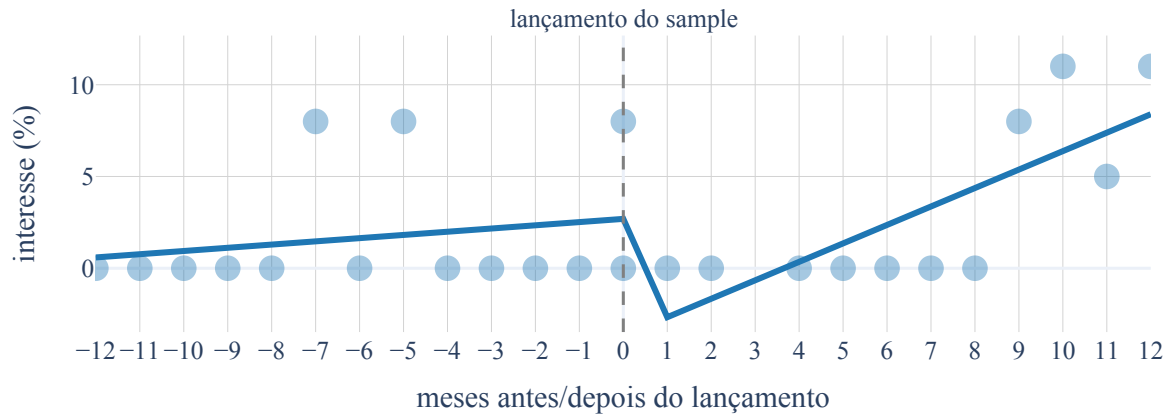


Figura B.21: Don't Stand So Close to Me (The Cure, 1985) sampleado em So Human (Lady Sovereign, 2009). ATE Normalizado: -5.471 — Causalidade de Granger: 0.54

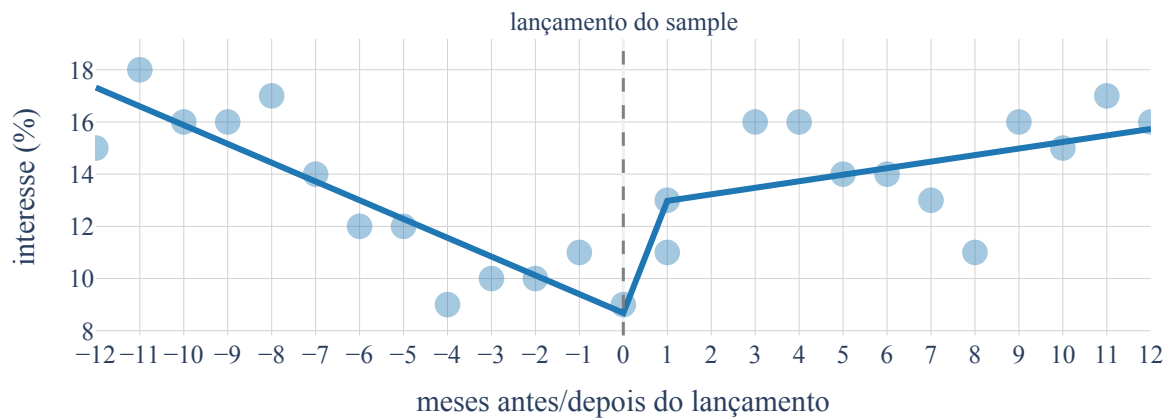


Figura B.22: Shots (LMFAO, 2009) sampleado em Somebody (Natalie La Rose, 2014). ATE Normalizado: 3.862 — Causalidade de Granger: 0.12

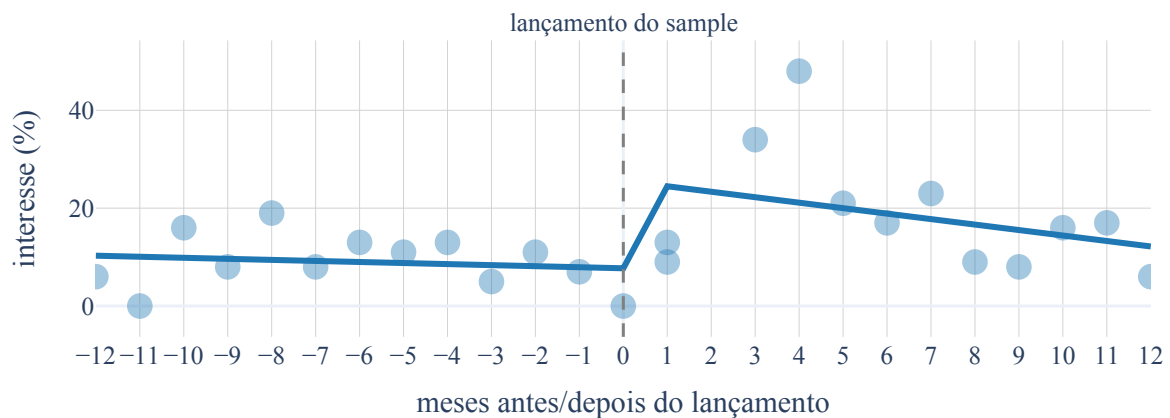


Figura B.23: Pon de Floor (Major Lazer, 2009) sampleado em Ass on the Floor (Diddy - Dirty Money, 2010). ATE Normalizado: 5.45 — Causalidade de Granger: 0.02

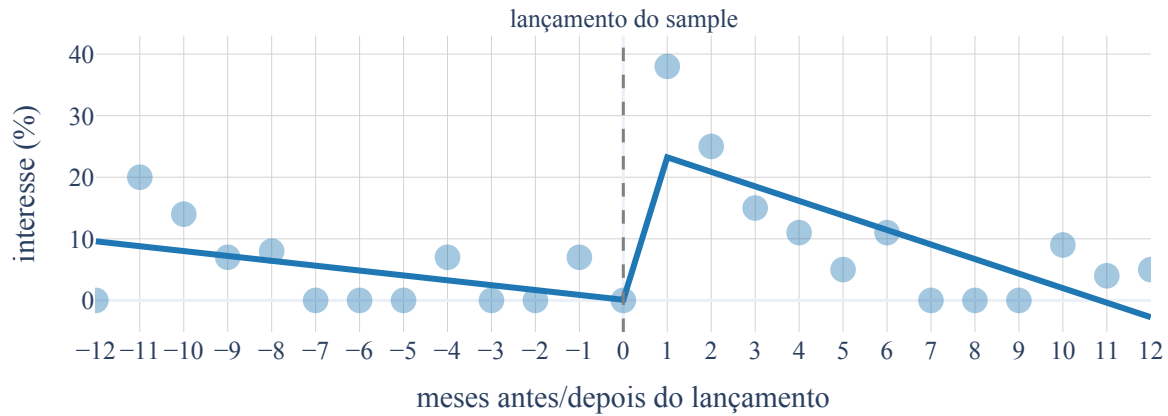


Figura B.24: Dirty Cash (Money Talks) (Adventures of Stevie V, 1989) sampleado em Dirtee Cash (Dizzee Rascal, 2009). ATE Normalizado: 10.157 — Causalidade de Granger: 0.34

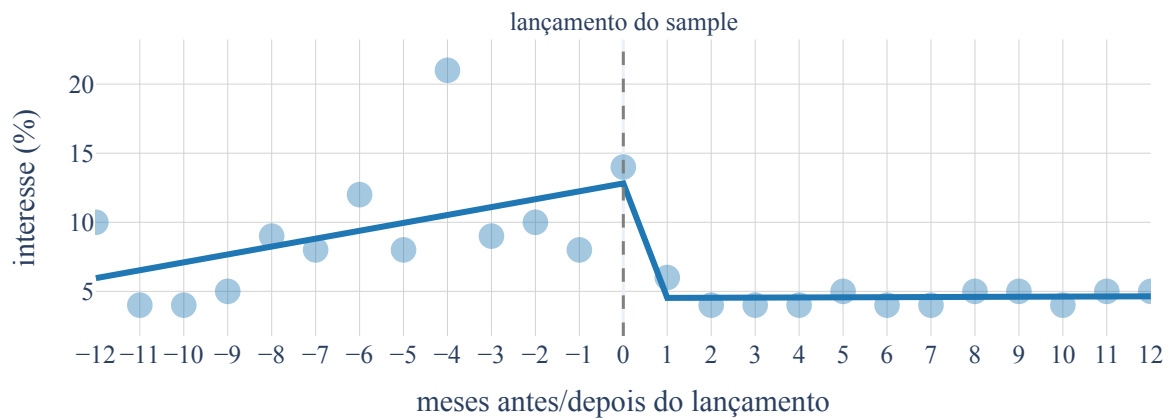


Figura B.25: (I Just) Died in Your Arms (Cutting Crew, 1986) sampleado em Relax, Take It Easy (Mika, 2006). ATE Normalizado: -4.186 — Causalidade de Granger: 0.3

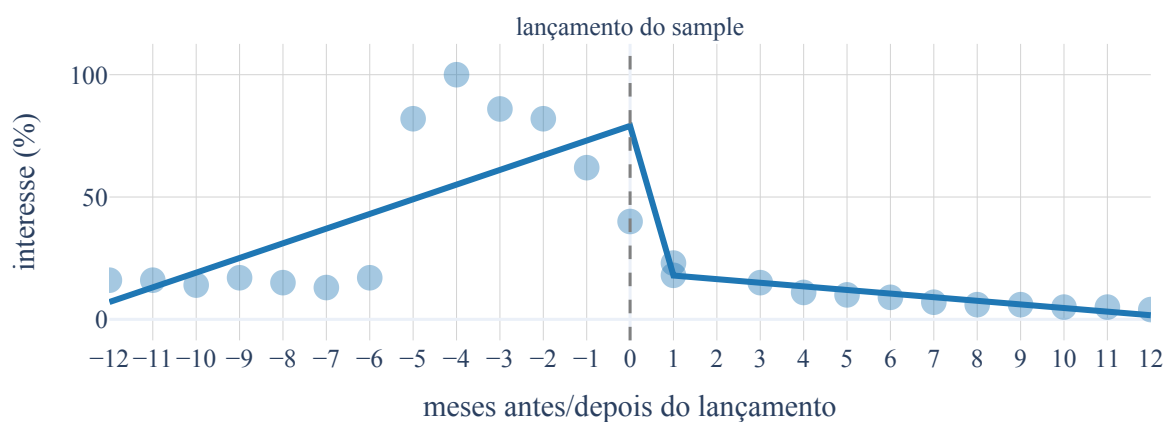


Figura B.26: Take It Off (Kesha, 2010) sampleado em Pop Culture (Madeon, 2011). ATE Normalizado: -4.337 — Causalidade de Granger: 0.0

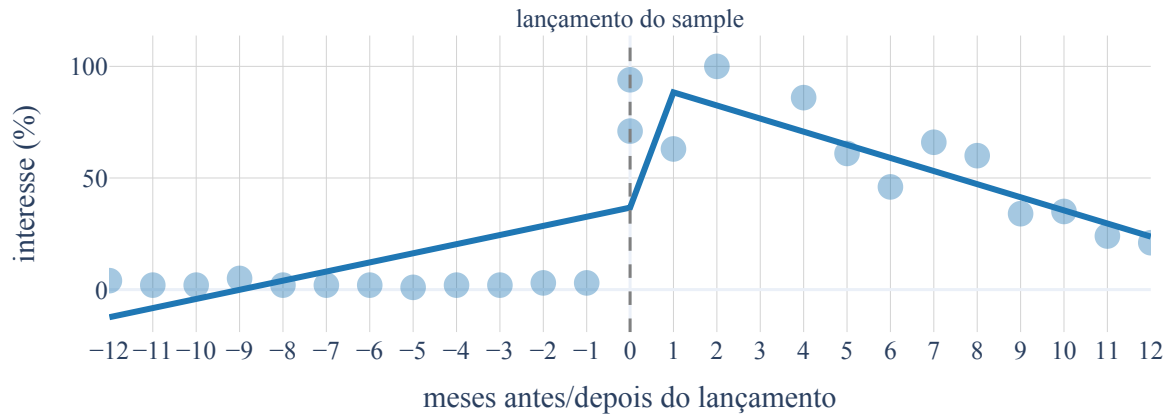


Figura B.27: One Way or Another (Blondie, 1978) regravado em One Way or Another (Teenage Kicks) (One Direction, 2013). ATE Normalizado: 5.06 — Causalidade de Granger: 0.93

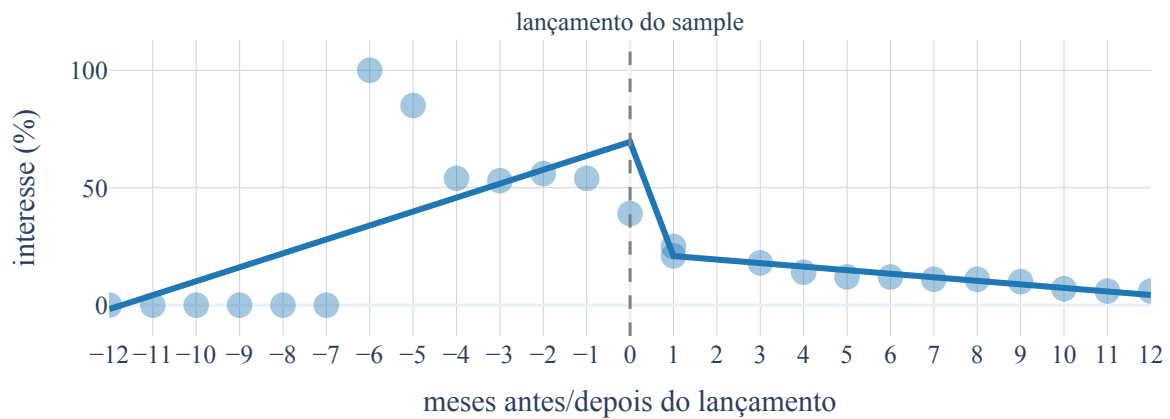


Figura B.28: Tom Ford (Jay-Z, 2013) sampleado em Dirty Vibe (Skrillex, 2014). ATE Normalizado: -4.23 — Causalidade de Granger: 0.48

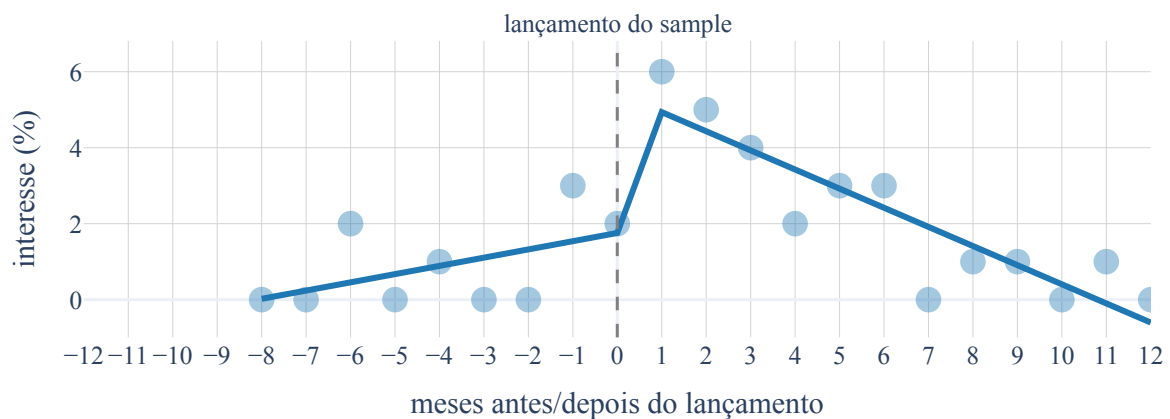


Figura B.29: Valerie (Steve Winwood, 1982) sampleado em Call on Me (Eric Prydz, 2004). ATE Normalizado: 5.351 — Causalidade de Granger: 0.6

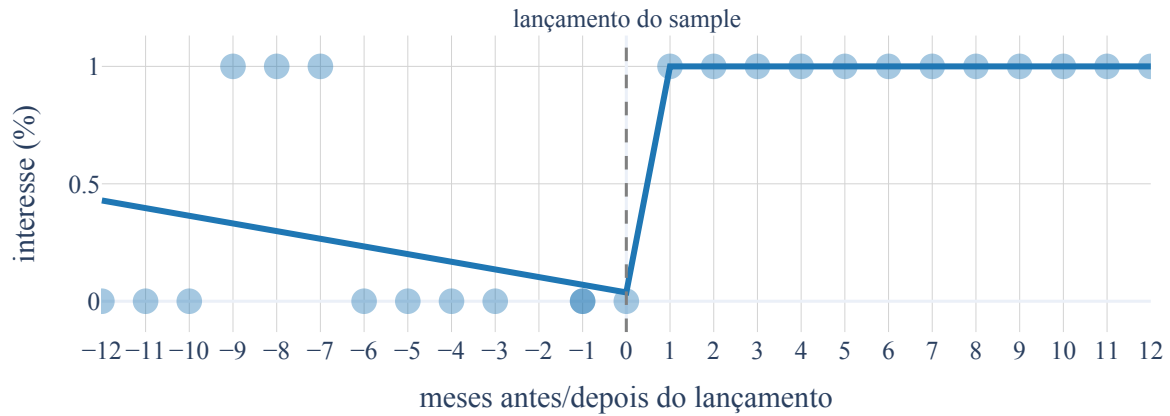


Figura B.30: I Won't Back Down (Tom Petty, 1989) sampleado em Stay with Me (Sam Smith, 2014). ATE Normalizado: 7.852 — Causalidade de Granger: 0.16

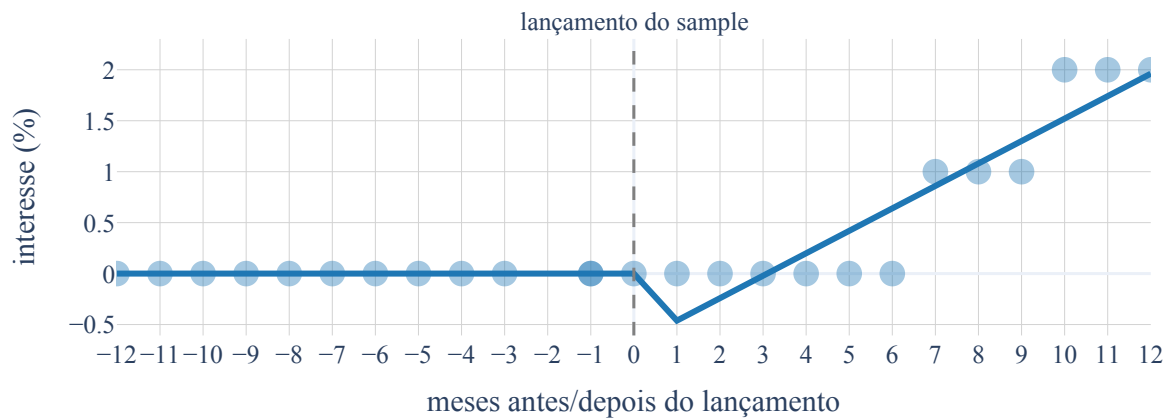


Figura B.31: Rocketeer (Far East Movement, 2010) sampleado em Somebody To Love (Justin Bieber, 2010). ATE Normalizado: -42.537 — Causalidade de Granger: 0.0

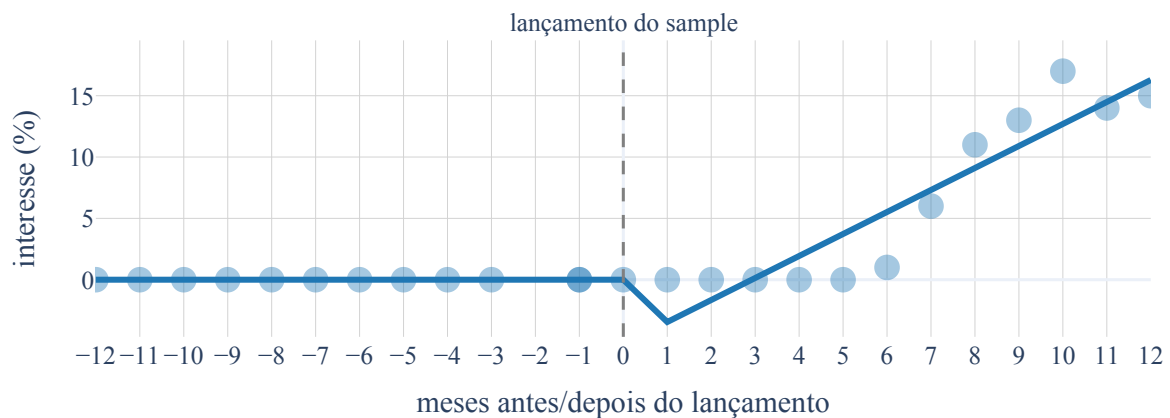


Figura B.32: Rocketeer (Far East Movement, 2010) sampleado em Like a G6 (Far East Movement, 2010). ATE Normalizado: -43.251 — Causalidade de Granger: 0.02

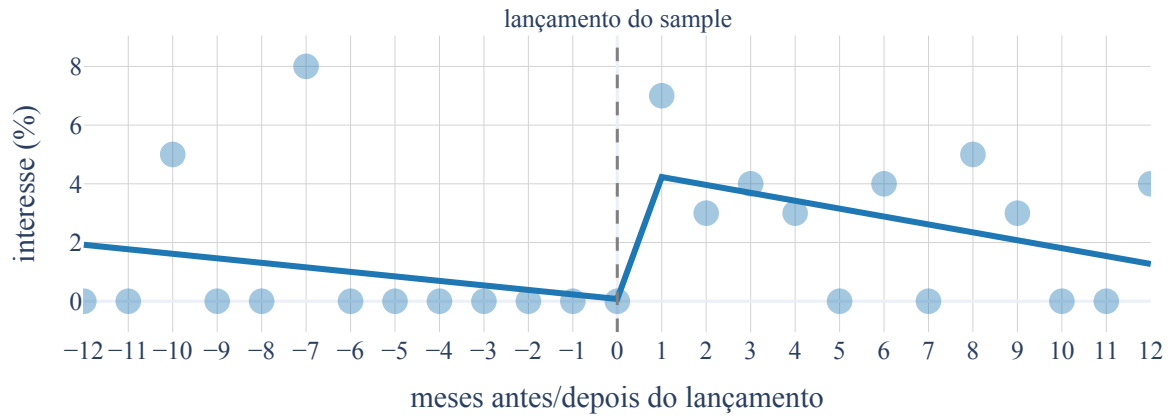


Figura B.33: N.Y. State of Mind (Nas, 1994) sampleado em Too Many Rappers (Beastie Boys, 2009). ATE Normalizado: 8.657 — Causalidade de Granger: 0.87

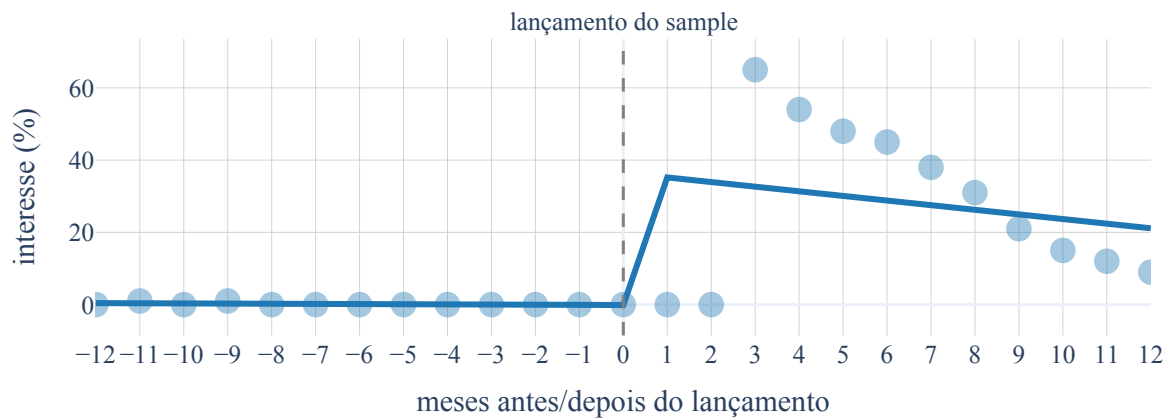


Figura B.34: The Motto (Drake, 2011) sampleado em She Will (Lil Wayne, 2011). ATE Normalizado: -10.413 — Causalidade de Granger: 0.03

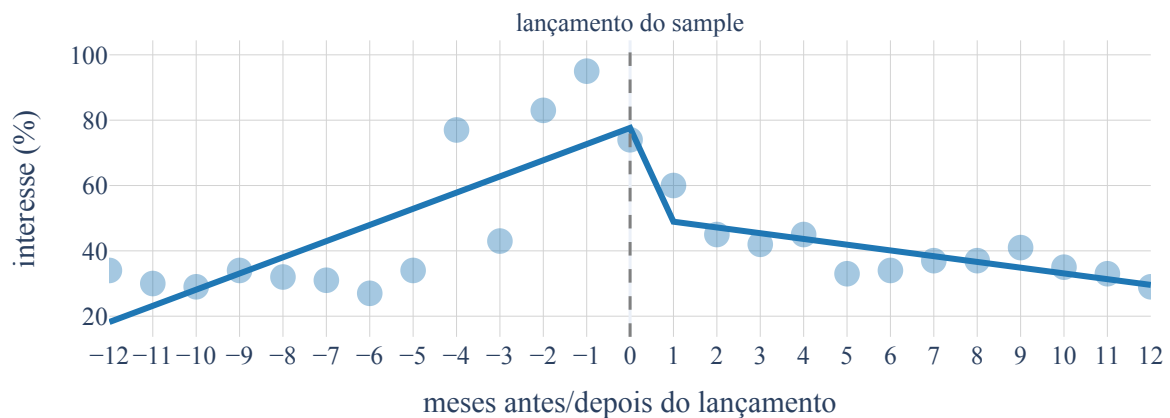


Figura B.35: Baby Got Back (Sir Mix-a-Lot, 1992) sampleado em Ew! (Jimmy Fallon, 2014). ATE Normalizado: -3.575 — Causalidade de Granger: 0.96

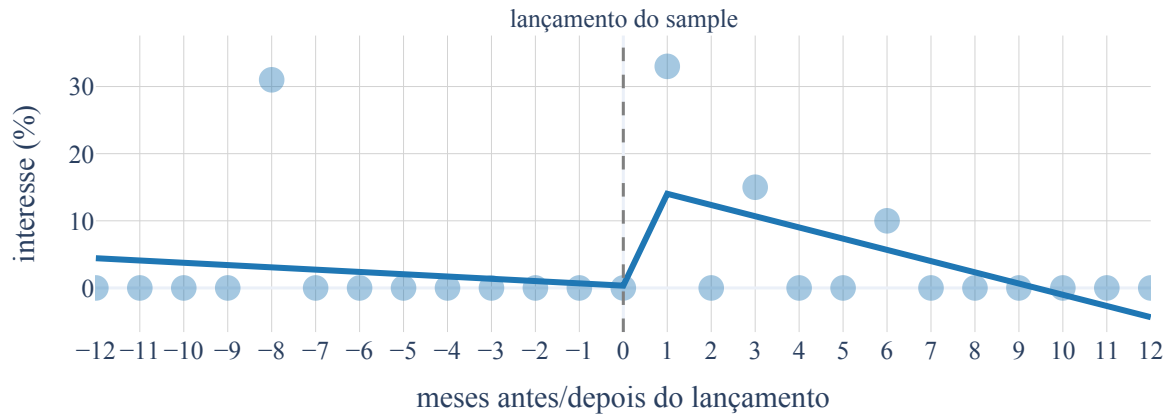


Figura B.36: Keep on Truckin' (Eddie Kendricks, 1973) sampleado em Do It Well (Jennifer Lopez, 2007). ATE Normalizado: 8.414 — Causalidade de Granger: 0.24

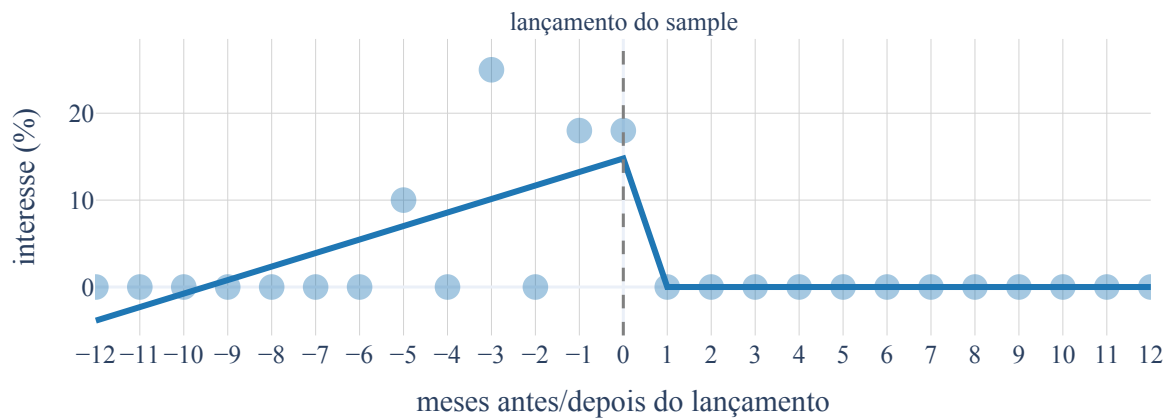


Figura B.37: Girls Can't Do What the Guys Do (Betty Wright, 1968) sampleado em Upgrade U (Beyoncé, 2006). ATE Normalizado: -4.615 — Causalidade de Granger: 0.95

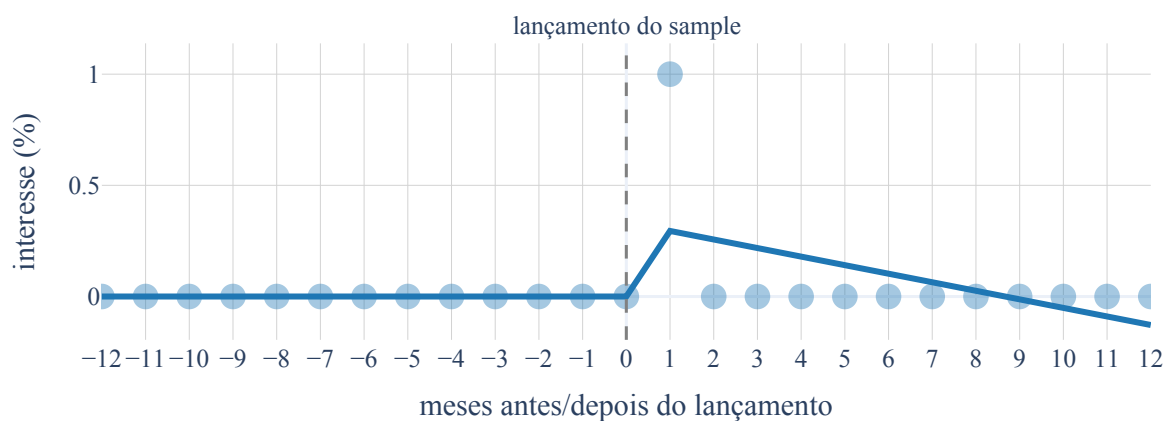


Figura B.38: Something's Got a Hold on Me (Etta James, 1962) sampleado em Good Feeling (Flo Rida, 2011). ATE Normalizado: 42.553 — Causalidade de Granger: 0.86

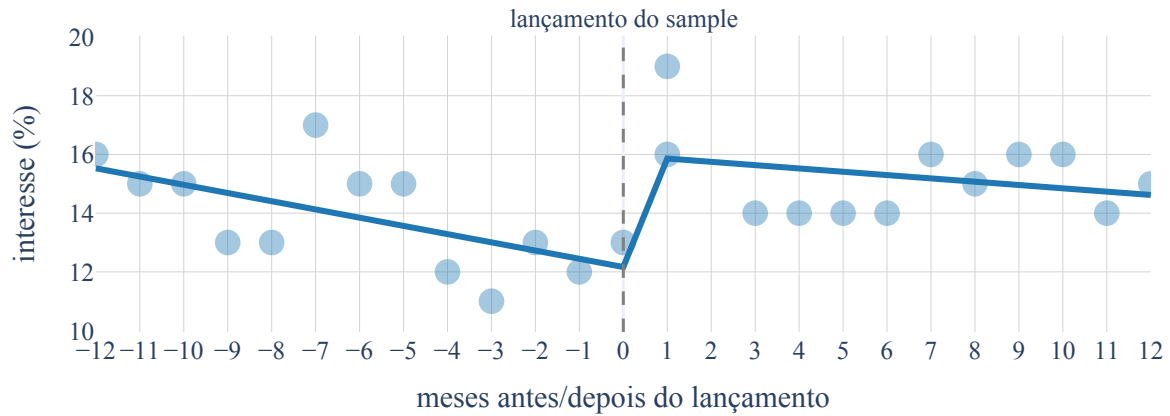


Figura B.39: Pony (Ginuwine, 1996) sampleado em Jump (Rihanna, 2012). ATE Normalizado: 3.475 — Causalidade de Granger: 0.71

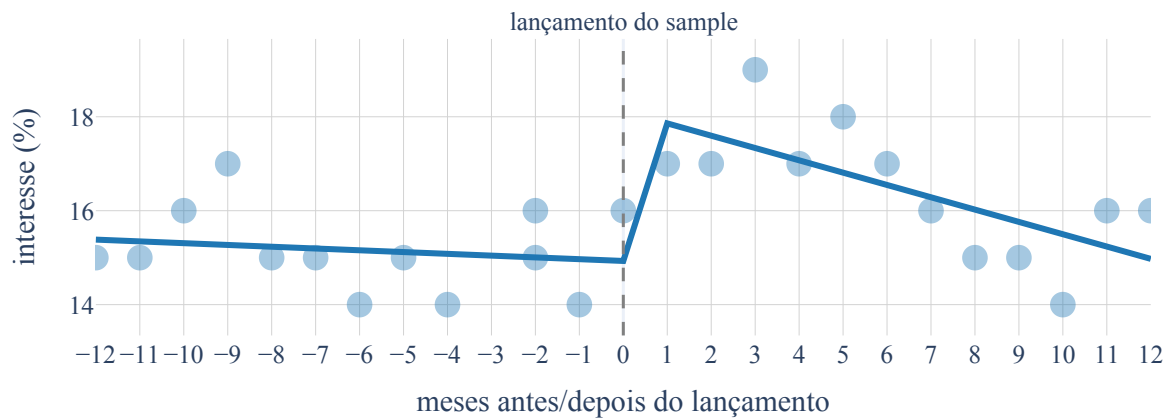


Figura B.40: I Will Survive (Gloria Gaynor, 1978) sampleado em "Hush Hush" (Pussycat Dolls, 2008). ATE Normalizado: 3.108 — Causalidade de Granger: 0.74

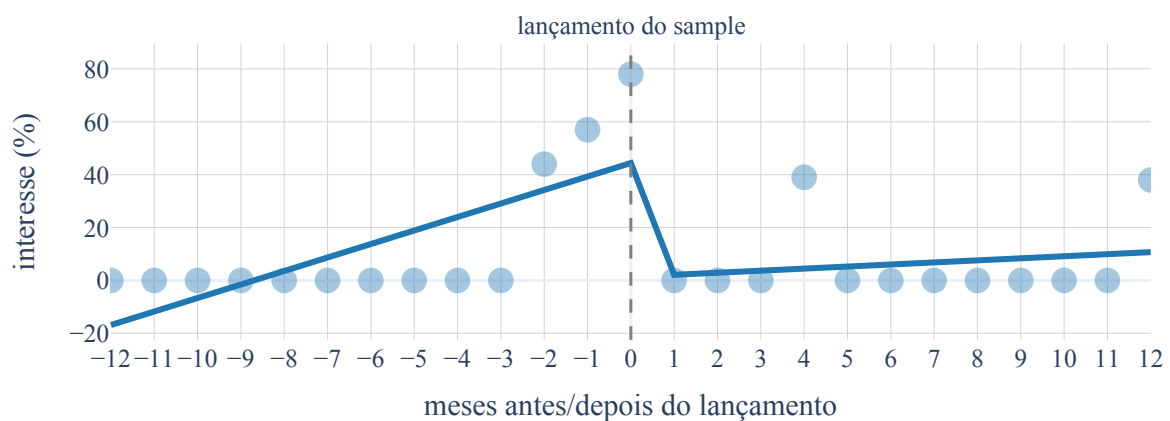


Figura B.41: Heartache Avenue (The Maisonettes, 1983) sampleado em The Avenue (Roll Deep, 2005). ATE Normalizado: -4.585 — Causalidade de Granger: 0.15

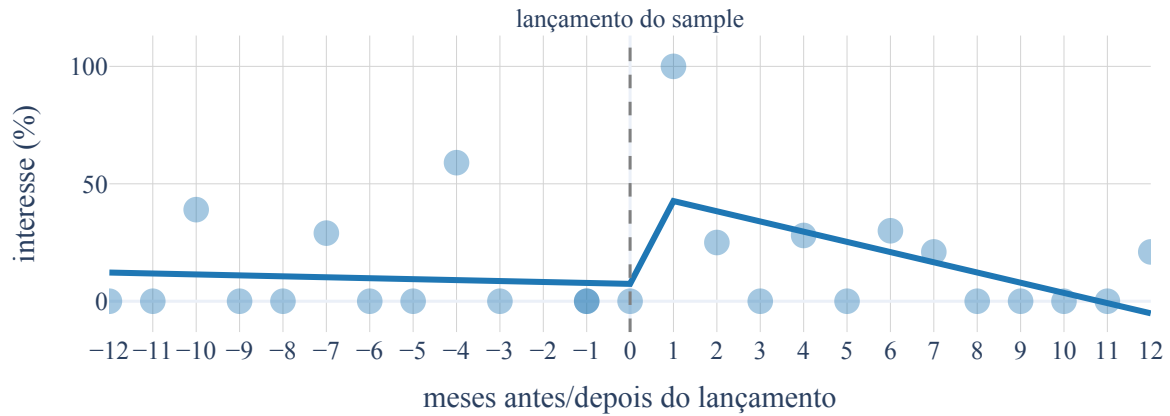


Figura B.42: Fade to Grey (Visage, 1980) sampleado em One Word (Kelly Osbourne, 2005). ATE Normalizado: 6.286 — Causalidade de Granger: 0.54

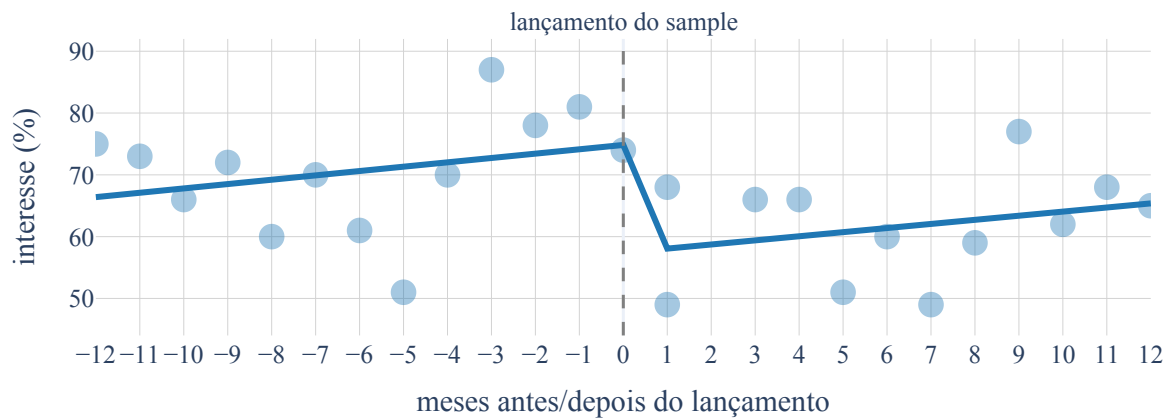


Figura B.43: Tocatta and Fugue in D minor, BWV 565 (Johann Sebastian Bach, 1707) sampleado em Baby's Coming Back/Transylvania (McFly, 2007). ATE Normalizado: -3.19 — Causalidade de Granger: 0.73

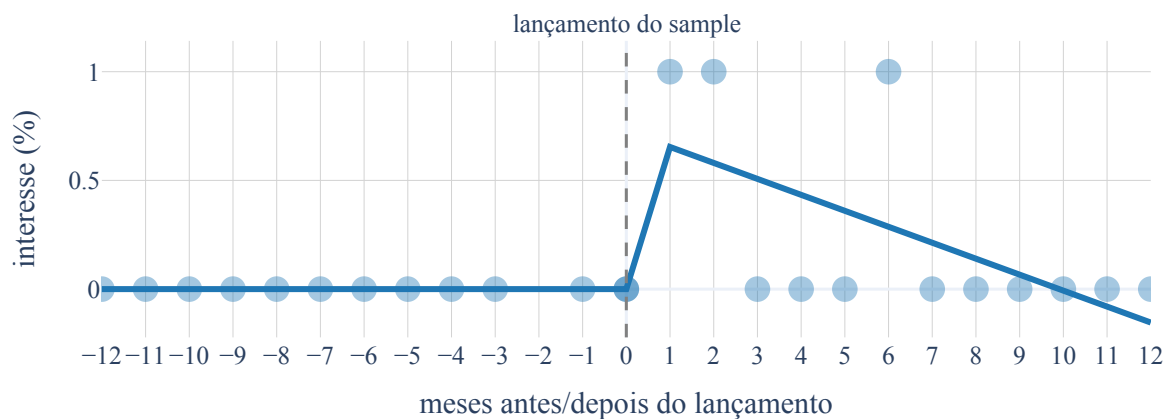


Figura B.44: One Nation Under a Groove (Funkadelic, 1978) sampleado em Peaches N Cream (Snoop Dogg song) (Snoop Dogg, 2015). ATE Normalizado: 39.051 — Causalidade de Granger: 0.88

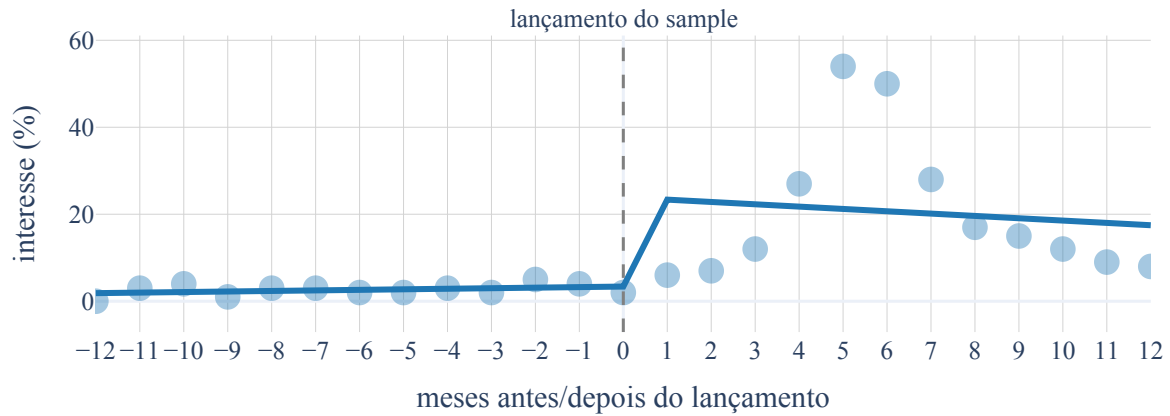


Figura B.45: Wanna Be Startin' Somethin' (Michael Jackson, 1982) sampleado em Don't Stop the Music (Rihanna, 2007). ATE Normalizado: 6.412 — Causalidade de Granger: 0.36

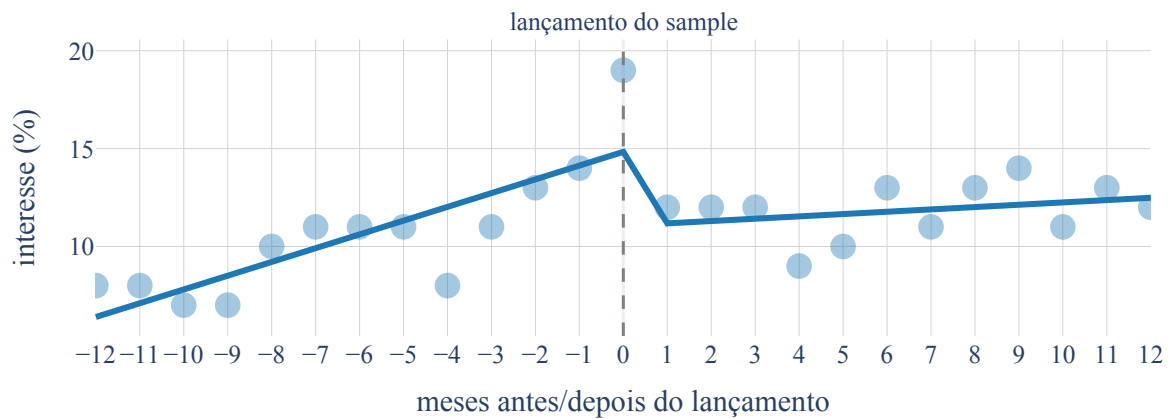


Figura B.46: No Woman, No Cry (Bob Marley, 1974) sampleado em Rayon de soleil (William Baldé, 2008). ATE Normalizado: -3.275 — Causalidade de Granger: 0.16

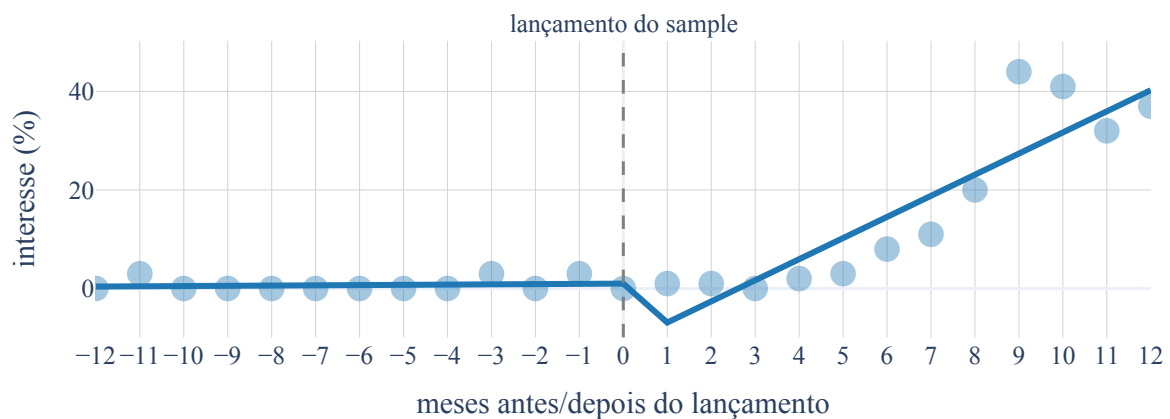


Figura B.47: Touch It (Busta Rhymes, 2005) sampleado em Technologic (Daft Punk, 2005). ATE Normalizado: -7.117 — Causalidade de Granger: 0.25

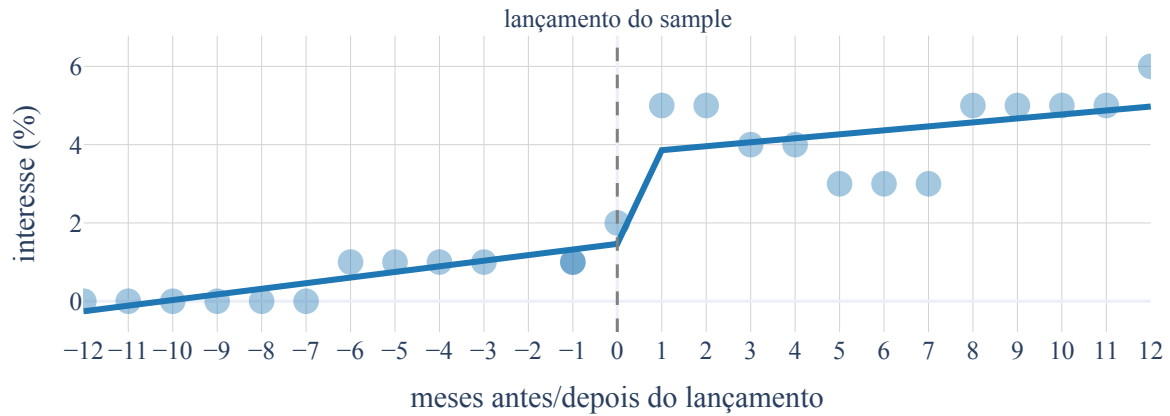


Figura B.48: Turn Down for What (DJ Snake, 2013) sampleado em Dynamite (Afrojack, 2014). ATE Normalizado: 5.059 — Causalidade de Granger: 1.0

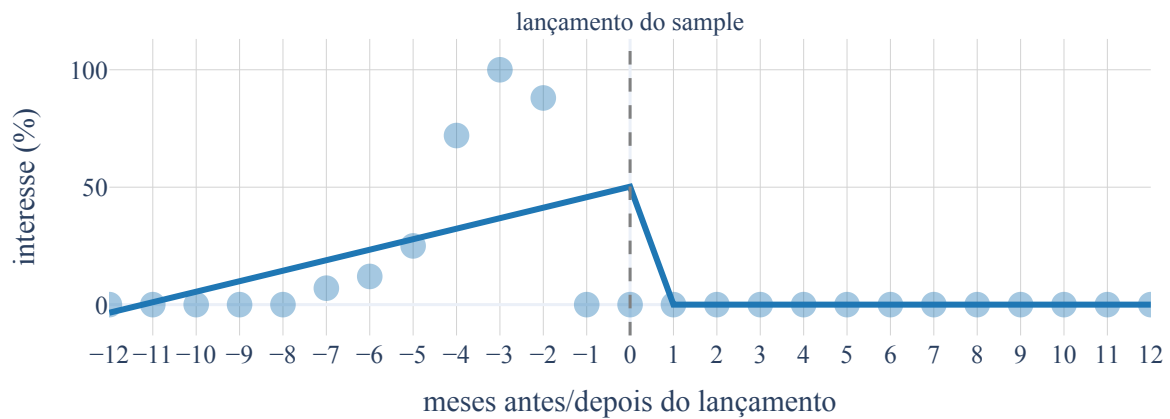


Figura B.49: Somebody To Love (Justin Bieber, 2010) sampleado em Rocketeer (Far East Movement, 2010). ATE Normalizado: -4.615 — Causalidade de Granger: 0.19

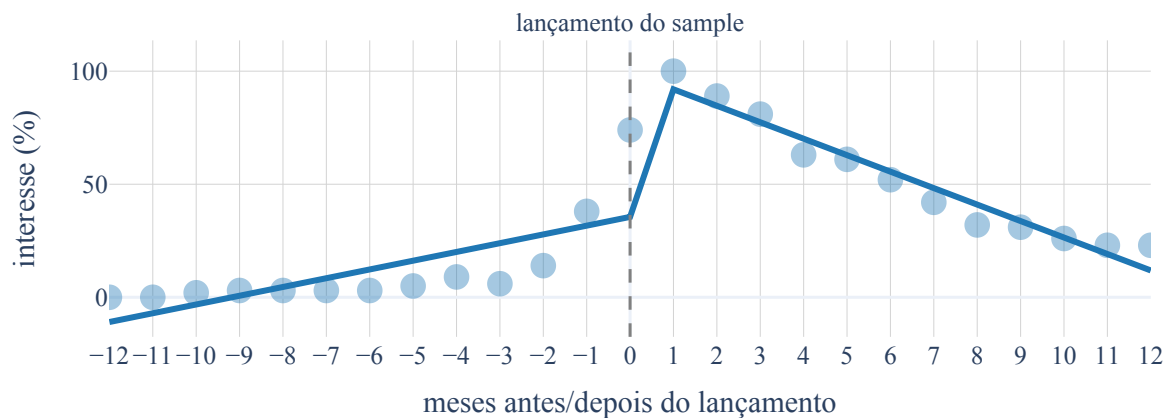


Figura B.50: Like a G6 (Far East Movement, 2010) sampleado em Rocketeer (Far East Movement, 2010). ATE Normalizado: 5.194 — Causalidade de Granger: 1.0

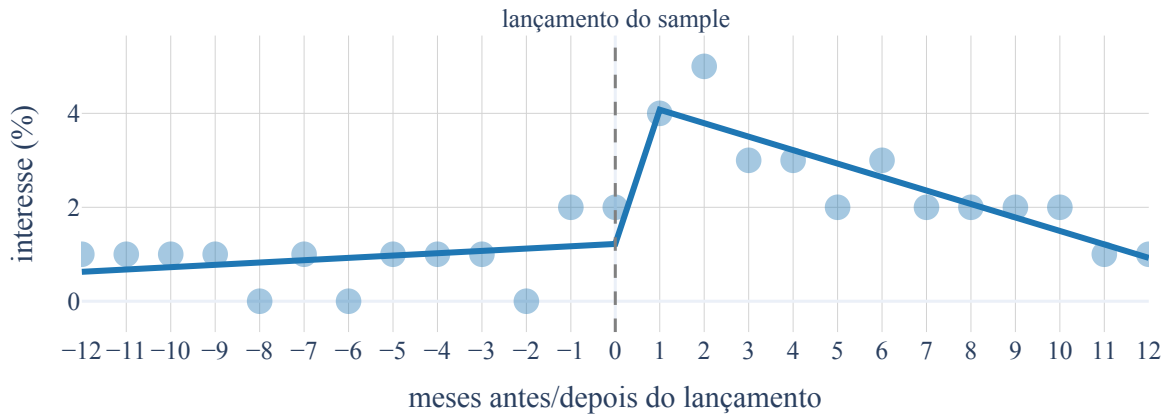


Figura B.51: The Rhythm of the Night (Corona, 1993) regravado em Of the Night (song) (Bastille, 2012). ATE Normalizado: 5.556 — Causalidade de Granger: 0.98

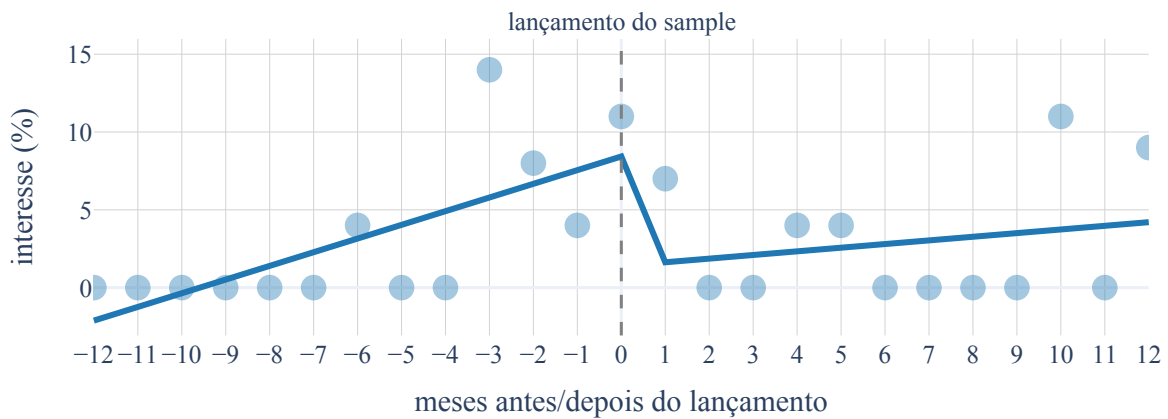


Figura B.52: In the Closet (Michael Jackson, 1991) sampleado em Take You Down (Chris Brown, 2007). ATE Normalizado: -4.436 — Causalidade de Granger: 0.06

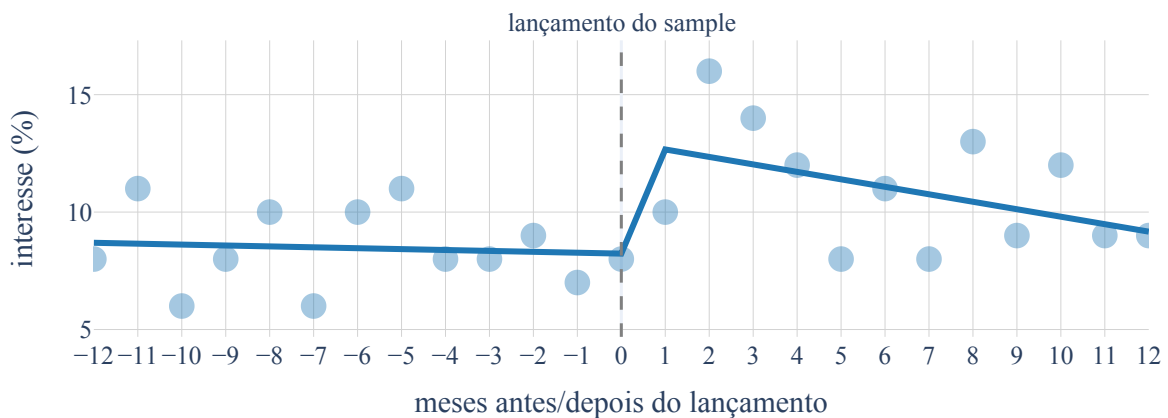


Figura B.53: Be My Lover (La Bouche, 1995) sampleado em Be My Lover (Inna, 2013). ATE Normalizado: 4.073 — Causalidade de Granger: 0.51

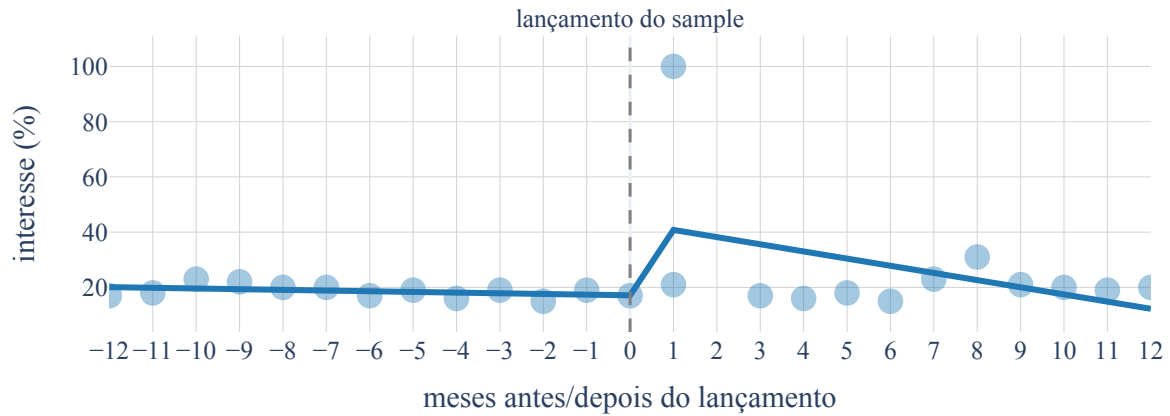


Figura B.54: Leck mich im Arsch (Wolfgang Amadeus Mozart, 1782) sampleado em Leck mich im Arsch (Insane Clown Posse, 2011). ATE Normalizado: 5.039 — Causalidade de Granger: 0.0

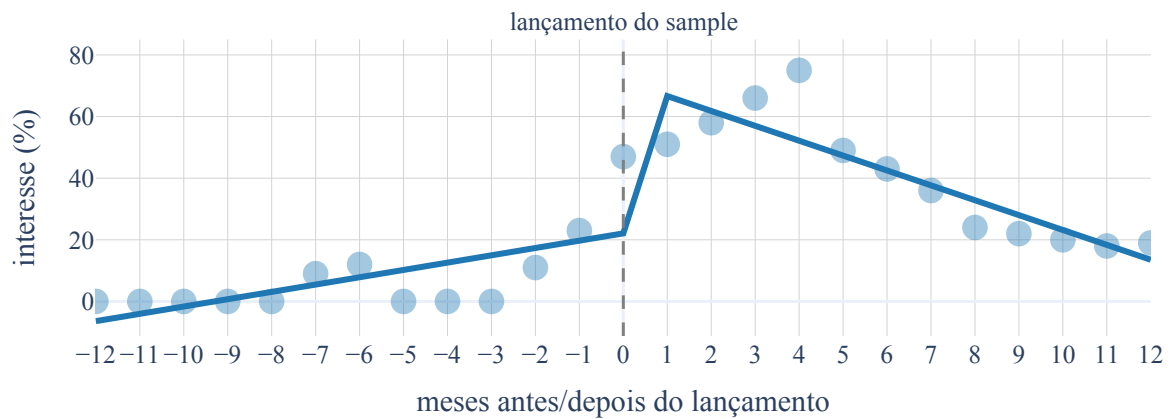


Figura B.55: Million Voices (Otto Knows, 2012) sampleado em 4 AM (Scooter, 2012). ATE Normalizado: 5.414 — Causalidade de Granger: 0.74

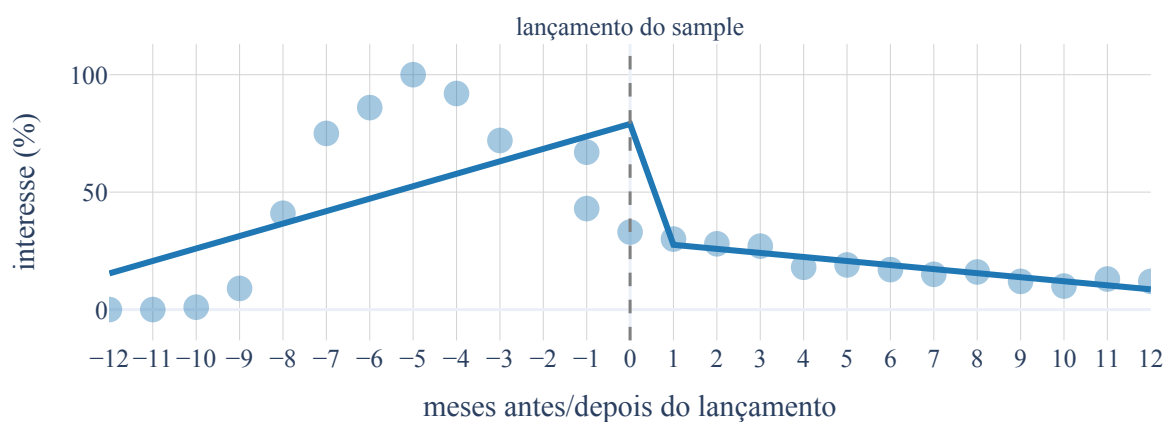


Figura B.56: Miss Independent (Ne-Yo, 2008) sampleado em Knock You Down (Keri Hilson, 2009). ATE Normalizado: -4.158 — Causalidade de Granger: 0.93

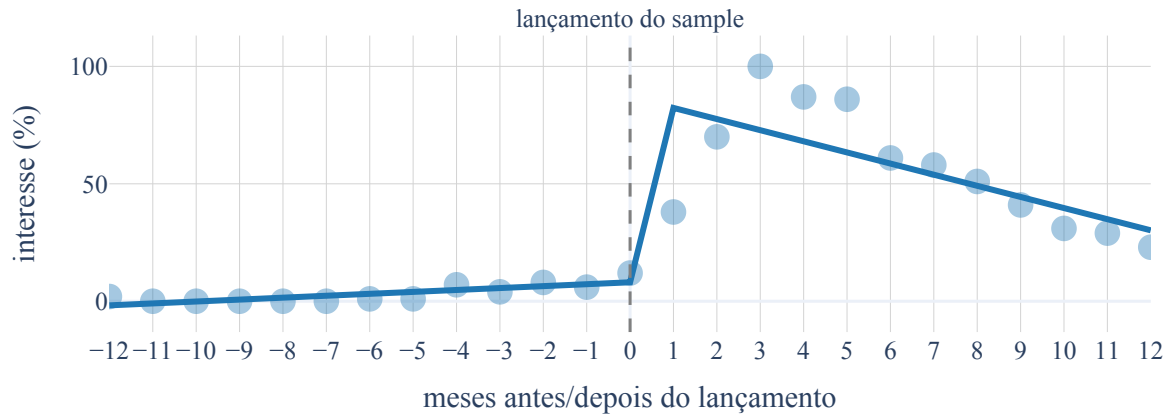


Figura B.57: Like a G6 (Far East Movement, 2010) sampleado em Booty Bounce (Dev, 2010). ATE Normalizado: 6.888 — Causalidade de Granger: 0.99

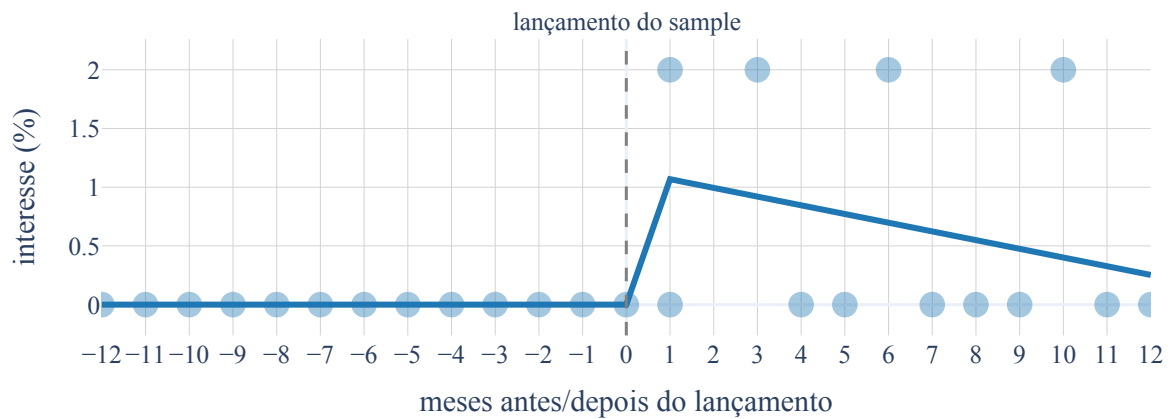


Figura B.58: Nightshift (Commodores, 1985) sampleado em Lose to Win (Fantasia, 2012). ATE Normalizado: -38.059 — Causalidade de Granger: 0.17

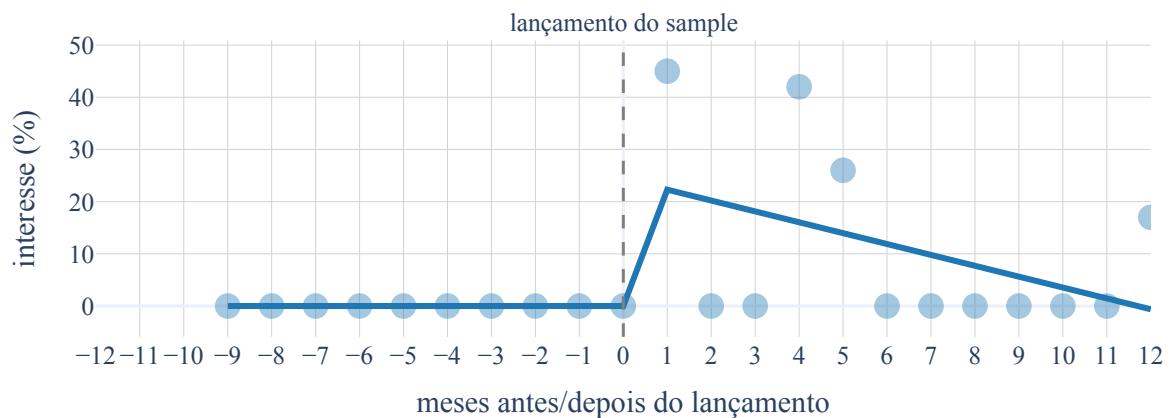


Figura B.59: Inseparable (Natalie Cole, 1975) sampleado em Karma (Lloyd Banks, 2004). ATE Normalizado: -40.092 — Causalidade de Granger: 0.0

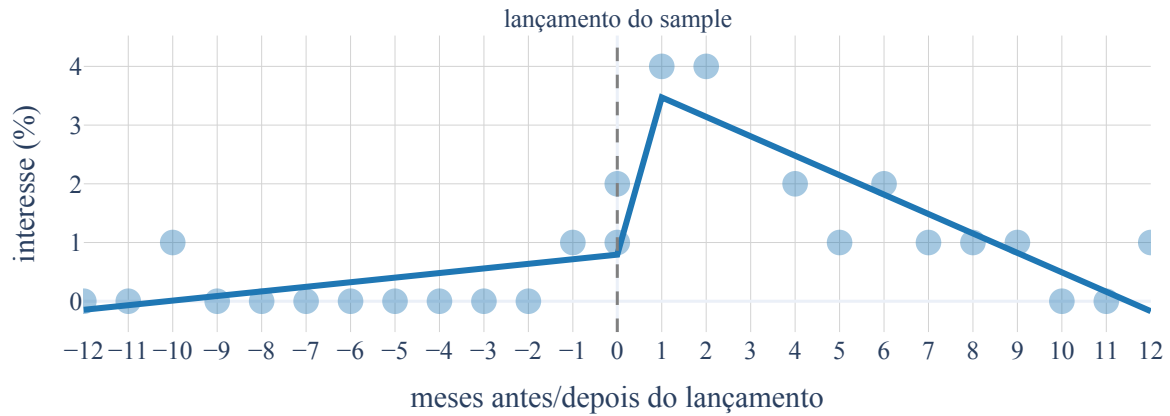


Figura B.60: Mr. Lonely (Bobby Vinton, 1964) sampleado em Lonely (Akon, 2004). ATE Normalizado: 5.941 — Causalidade de Granger: 0.72

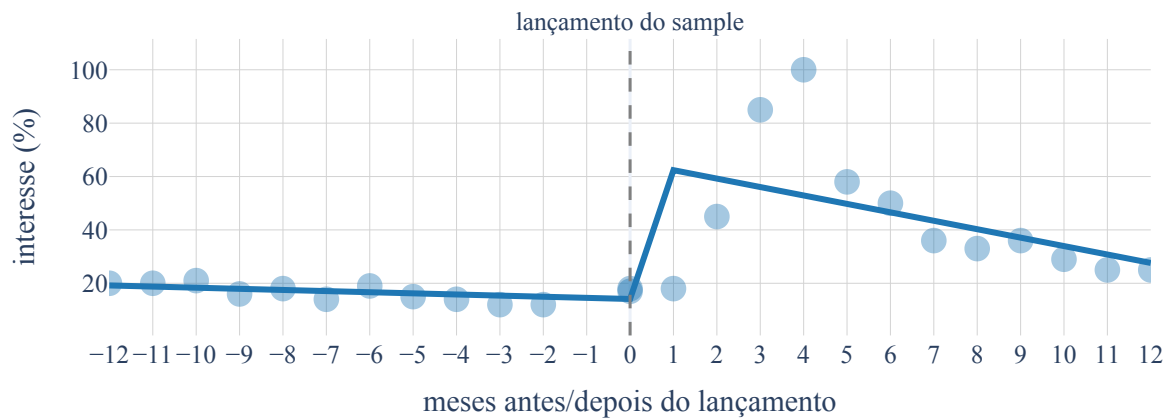


Figura B.61: Party & Bullshit (The Notorious B.I.G., 1993) sampleado em How We Do (Party) (Rita Ora, 2012). ATE Normalizado: 5.901 — Causalidade de Granger: 0.12

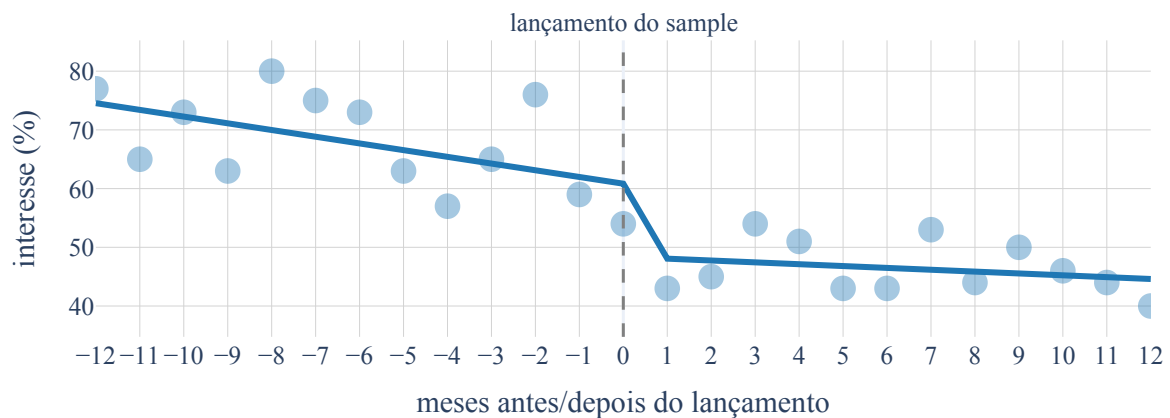


Figura B.62: Stronger (Britney Spears, 2000) sampleado em Pound the Alarm (Nicki Minaj, 2012). ATE Normalizado: -3.067 — Causalidade de Granger: 0.04

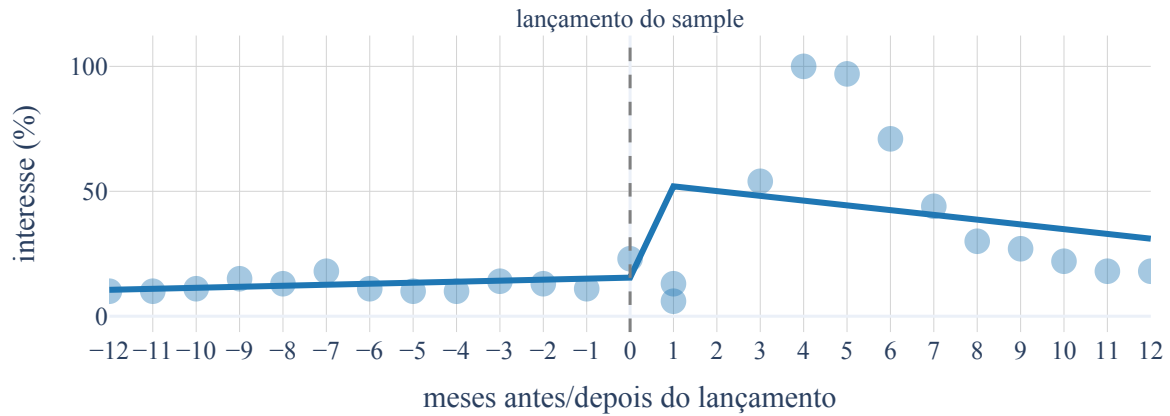


Figura B.63: Na Na Hey Hey Kiss Him Goodbye (Steam, 1969) sampleado em Goodbye (Kristinia Debarge, 2009). ATE Normalizado: 5.522 — Causalidade de Granger: 0.94

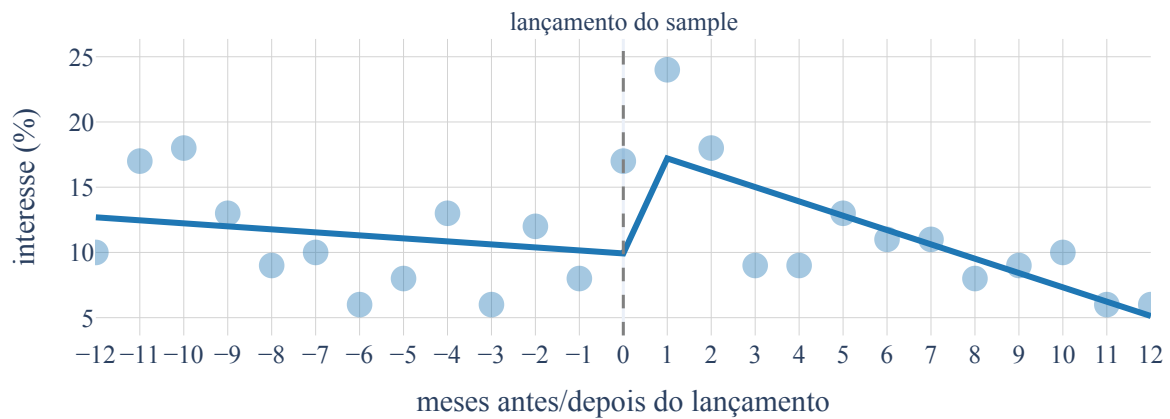


Figura B.64: My Cherie Amour (Stevie Wonder, 1969) sampleado em Wish You Were Mine (Philip George, 2014). ATE Normalizado: 4.448 — Causalidade de Granger: 0.01

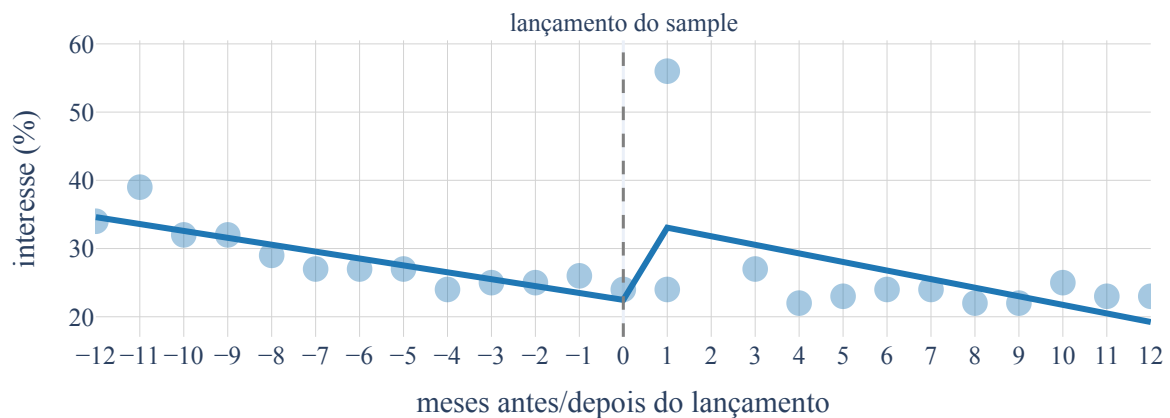


Figura B.65: Baba O'Riley (The Who, 1971) sampleado em Escape Velocity (The Chemical Brothers, 2010). ATE Normalizado: 3.984 — Causalidade de Granger: 0.13

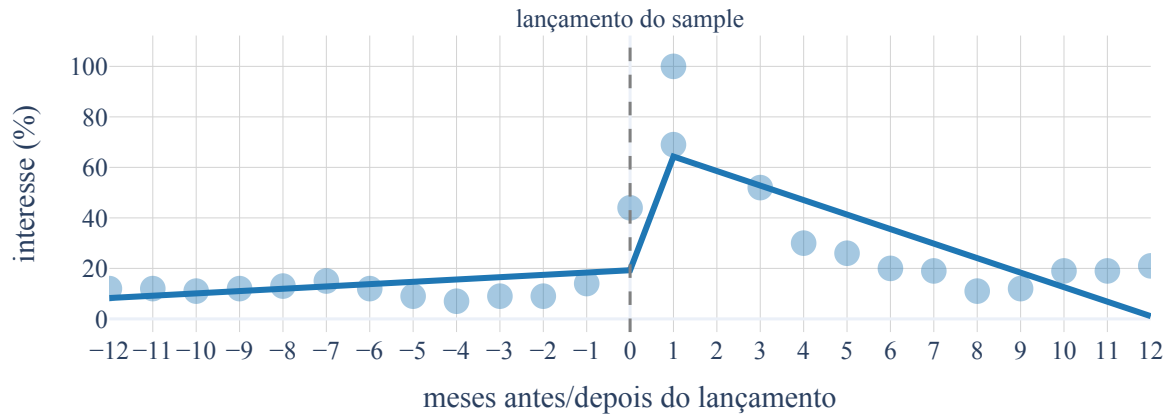


Figura B.66: You Spin Me Round (Like a Record) (Dead or Alive, 1984) sampleado em Right Round (Flo Rida, 2009). ATE Normalizado: 5.577 — Causalidade de Granger: 0.28

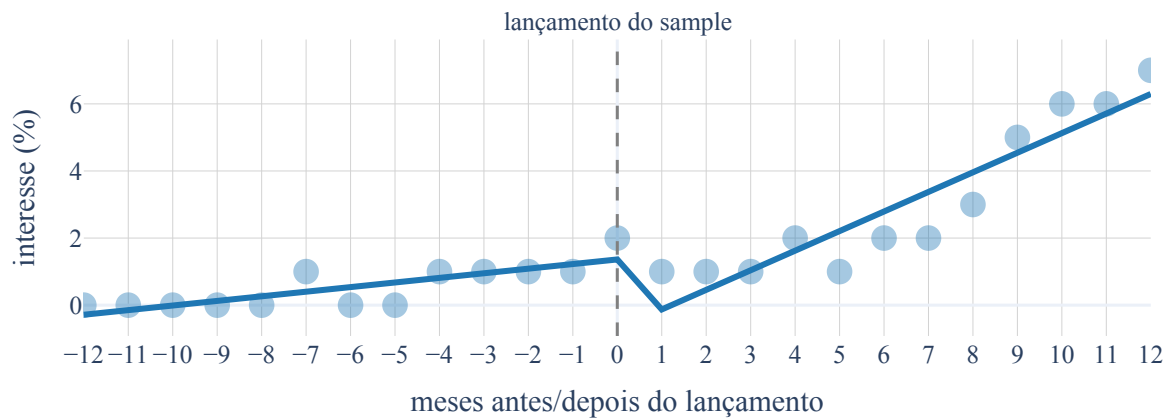


Figura B.67: Summertime Sadness (Lana Del Rey, 2012) sampleado em Body Electric (Lana Del Rey, 2012). ATE Normalizado: -5.032 — Causalidade de Granger: 0.47

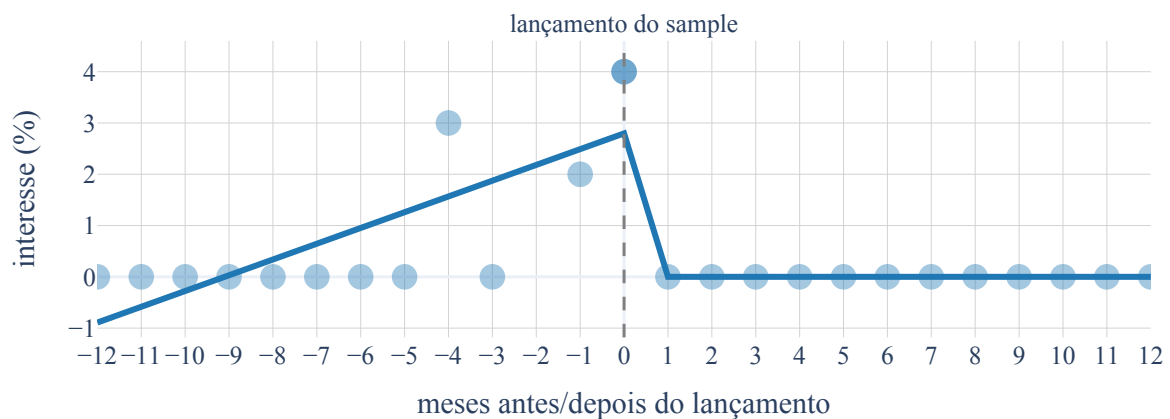


Figura B.68: Fader (The Temper Trap, 2009) remixado em Fade (Jakwob song) (Jakwob, 2009). ATE Normalizado: -4.615 — Causalidade de Granger: 0.51

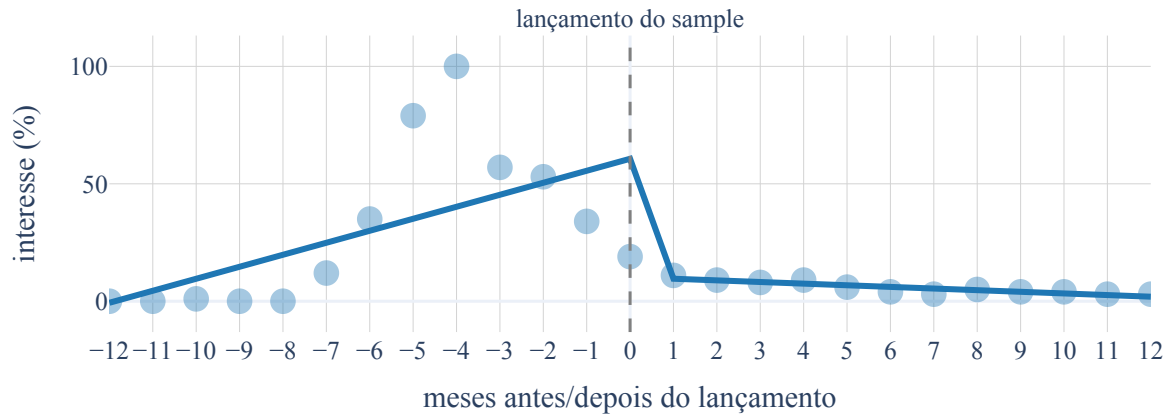


Figura B.69: Wearing My Rolex (Wiley, 2008) sampleado em Rolex Sweep (Skeptta, 2008). ATE Normalizado: -4.432 — Causalidade de Granger: 0.99

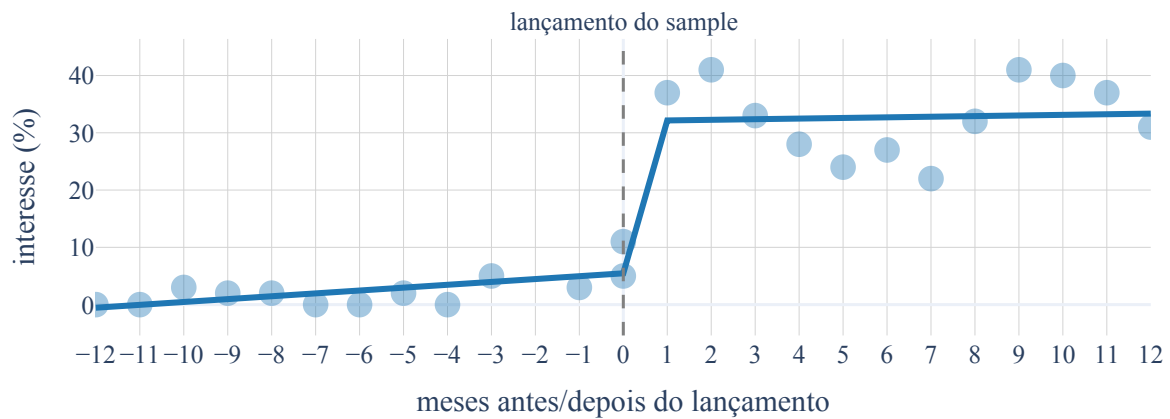


Figura B.70: Calabria (Rune, 2003) sampleado em Destination Calabria (Alex Gaudino, 2006). ATE Normalizado: 6.189 — Causalidade de Granger: 0.0

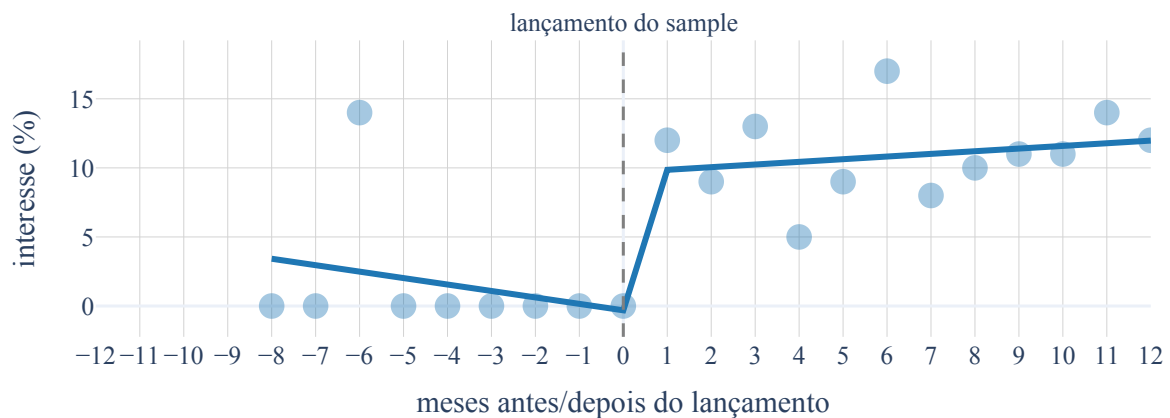


Figura B.71: Homecoming (Green Day, 2004) sampleado em American Idiot (Green Day, 2004). ATE Normalizado: -8.073 — Causalidade de Granger: 0.74

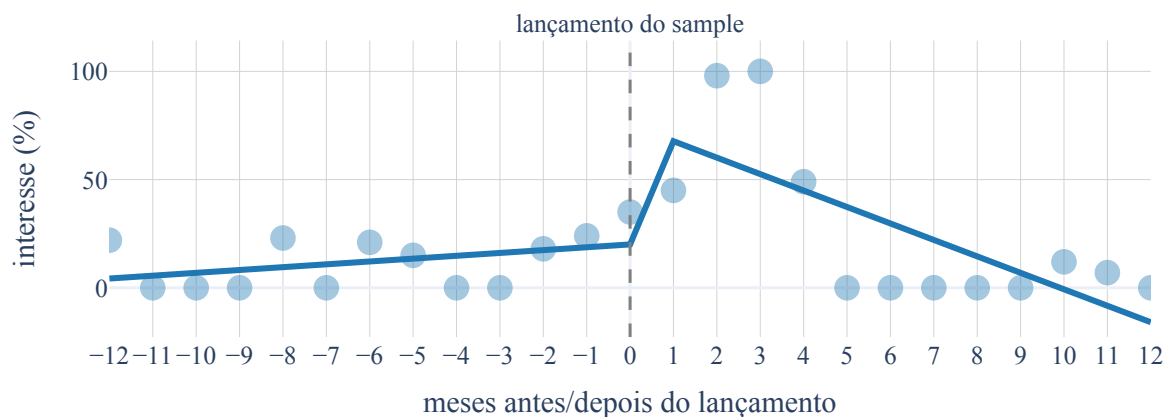


Figura B.72: We Don't Have to Take Our Clothes Off (Jermaine Stewart, 1986) sampleado em Clothes Off!! (Gym Class Heroes, 2007). ATE Normalizado: 5.624 — Causalidade de Granger: 0.95

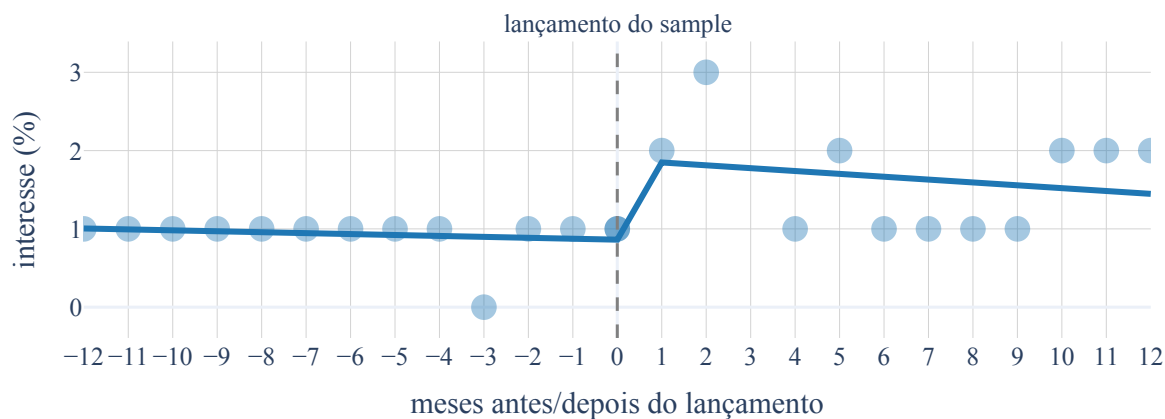


Figura B.73: Again (Janet Jackson, 1993) sampleado em Solo (Iyaz, 2010). ATE Normalizado: 4.785 — Causalidade de Granger: 0.14

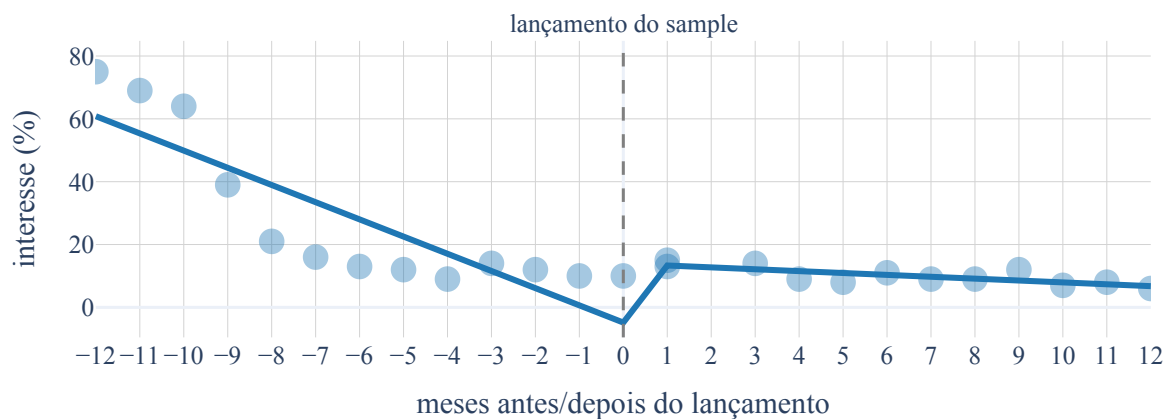


Figura B.74: Let It Rock (Kevin Rudolf, 2008) sampleado em I Made It (Cash Money Heroes) (Kevin Rudolf, 2010). ATE Normalizado: -5.963 — Causalidade de Granger: 0.07

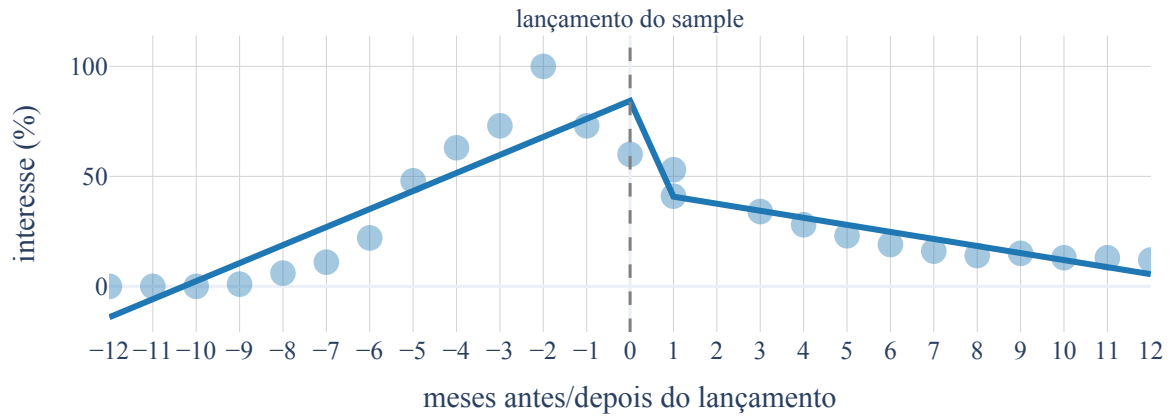


Figura B.75: Down (Jay Sean, 2009) sampleado em I Made It (Cash Money Heroes) (Kevin Rudolf, 2010). ATE Normalizado: -3.889 — Causalidade de Granger: 0.41

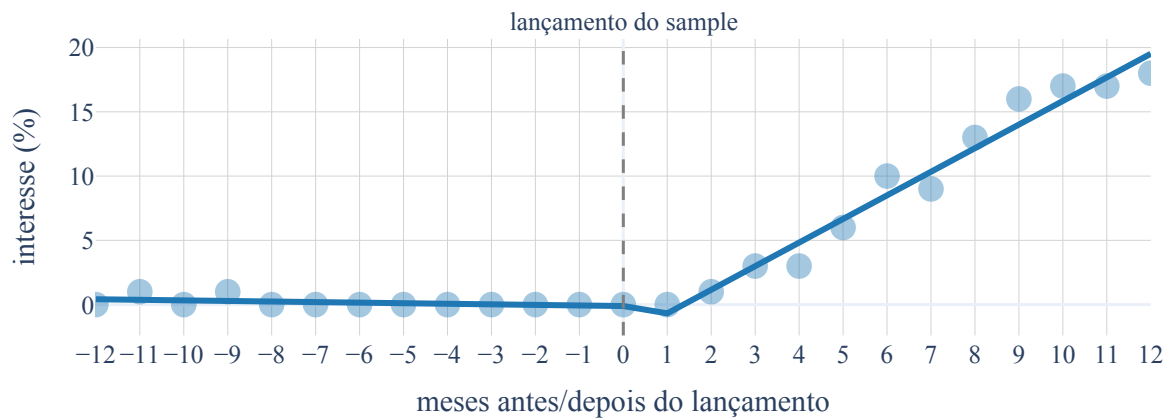


Figura B.76: Ima Korean (Rucka Rucka Ali, 2009) regravado em I Gotta Feeling (Black Eyed Peas, 2009). ATE Normalizado: 7.692 — Causalidade de Granger: 0.02

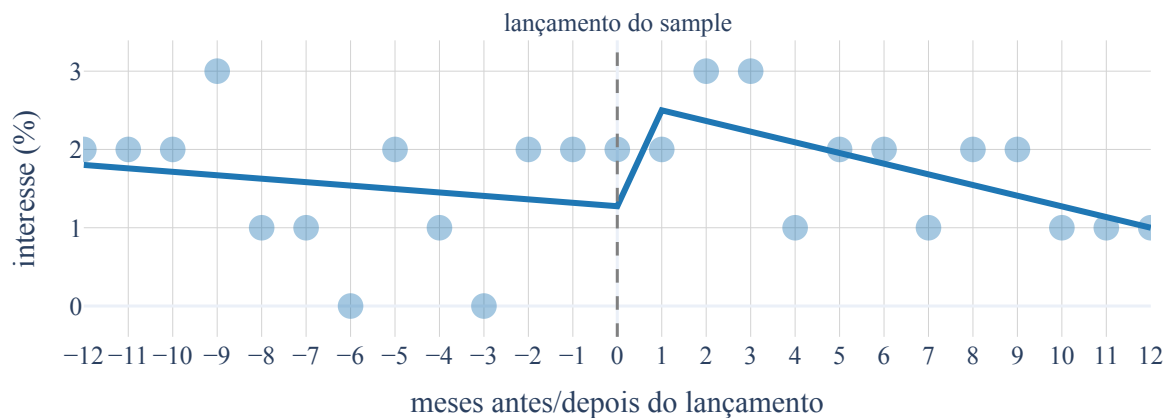


Figura B.77: Heaven Is a Place on Earth (Belinda Carlisle, 1987) regravado em Heaven Is a Place on Earth (Wildside, 2009). ATE Normalizado: 4.68 — Causalidade de Granger: 0.0

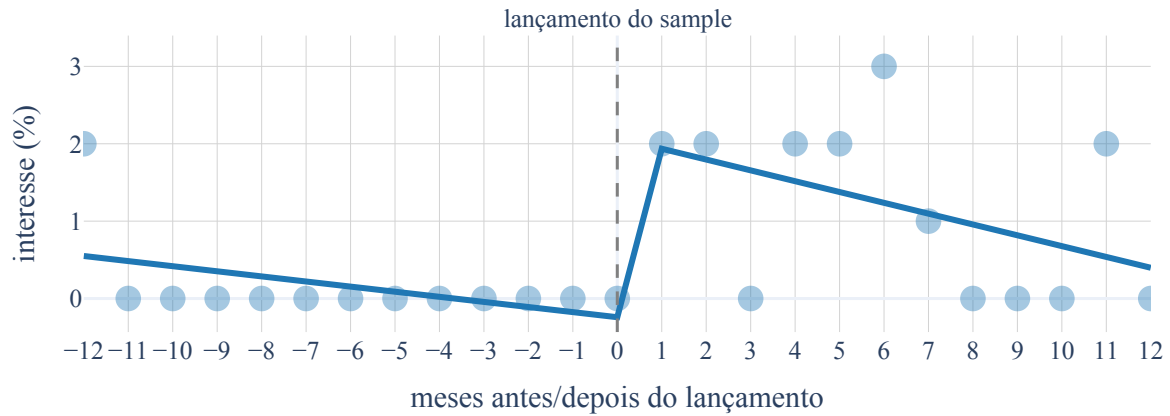


Figura B.78: She's on Fire (Amy Holland, 1983) sampleado em On Fire (Lil Wayne, 2009). ATE Normalizado: -6.867 — Causalidade de Granger: 0.58

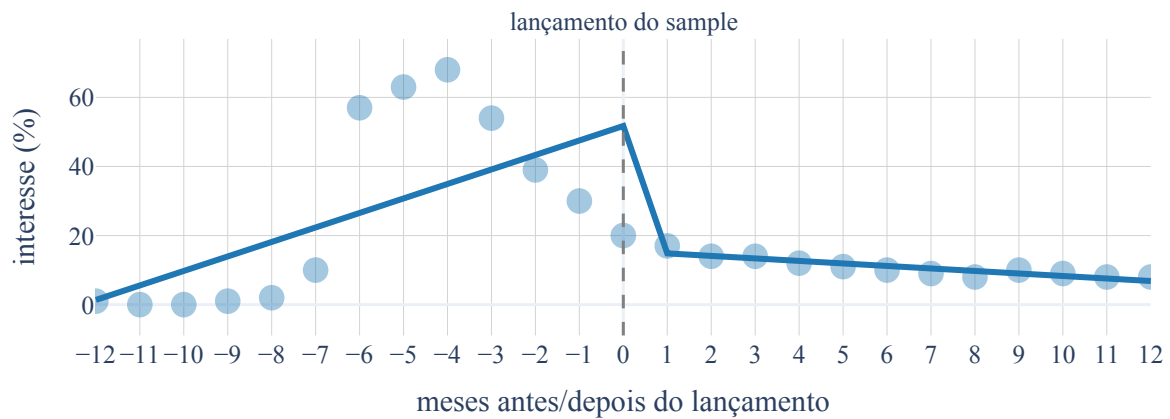


Figura B.79: The One That Got Away (Katy Perry, 2010) regravado em The One That Got Away (Jake Coco, 2011). ATE Normalizado: -4.262 — Causalidade de Granger: 0.89

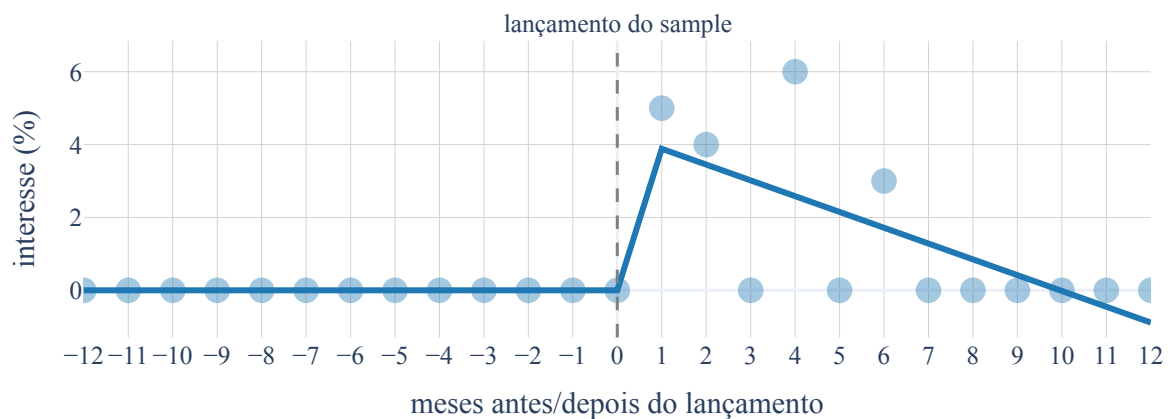


Figura B.80: Been Down So Long (The Doors, 1971) sampleado em Hypest Hype (Chase & Status, 2010). ATE Normalizado: 41.03 — Causalidade de Granger: 0.93

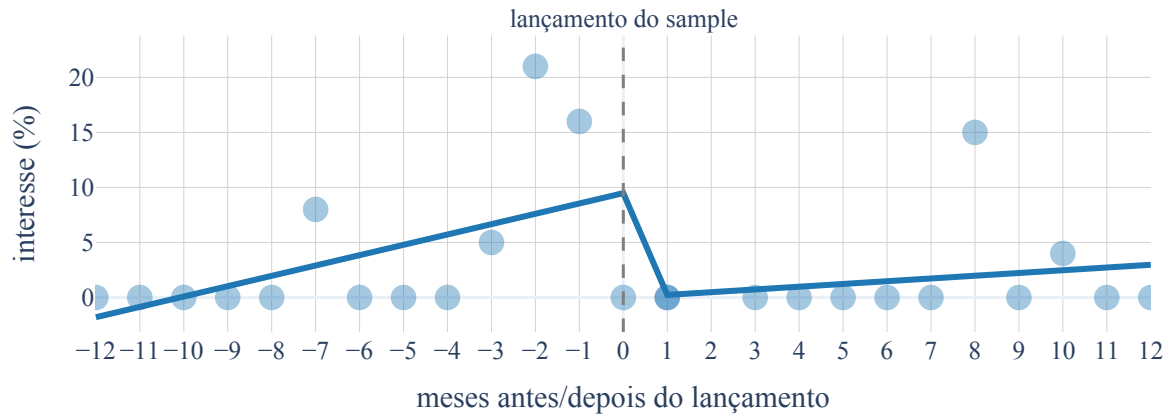


Figura B.81: Be Happy (Mary J. Blige, 1994) sampleado em Holding You Down (Goin' in Circles) (Jazmine Sullivan, 2010). ATE Normalizado: -4.617 — Causalidade de Granger: 0.1

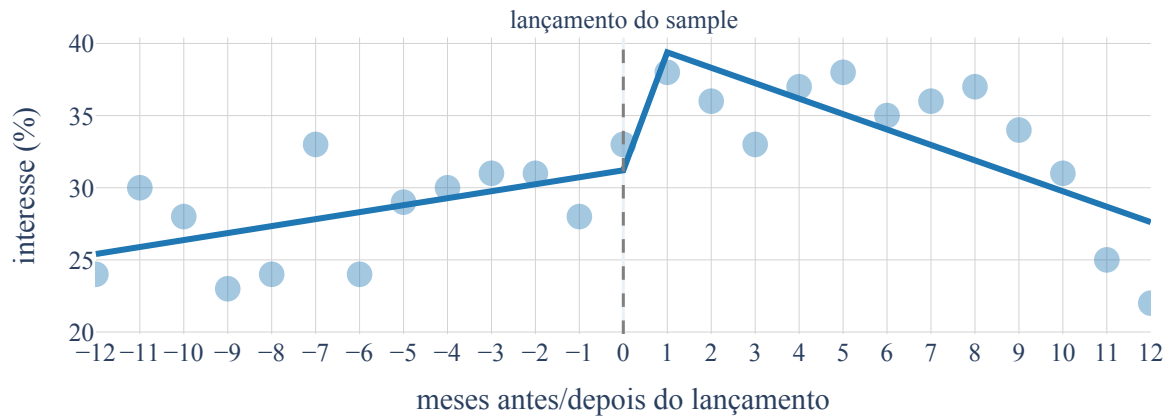


Figura B.82: Music Sounds Better with You (Stardust, 1998) sampleado em Music Sounds Better with U (Big Time Rush, 2011). ATE Normalizado: 3.422 — Causalidade de Granger: 0.9