

O Que os Países Escutam: Analisando a Rede de Gêneros Musicais ao Redor do Mundo.

Title: What Countries Listen to: Analyzing the Network of Music Genres around the World

Maria Luiza Botelho Mondelli¹, Luiz M. R. Gadelha Jr.¹, Artur Ziviani¹

¹Laboratório Nacional de Computação Científica (LNCC)
Petrópolis, Rio de Janeiro – Brasil

{mluiza, lgadelha, ziviani}@lncc.br

Abstract. *Music streaming platforms are increasingly popular, facilitating the access to music content. This effect extends the reach of different musical styles, increasing the diversity of listened music genres in different countries around the world. In order to better understand this diversity, in this paper we build and analyze a complex network of artists, music genres, and countries using data from Spotify. As a result, in addition to identifying communities of countries with similar music genres, we show how the diversity of music genres can influence the modeling and analysis of the considered network. We also classify the most commonly listened genres using centrality metrics and we analyze how the diffusion of genres occurs over time, including a case study with a viral music.*

Keywords. *Complex networks; Network science; Music popularity.*

Resumo. *Plataformas de streaming de música são cada vez mais populares, facilitando o acesso ao conteúdo musical. Esse efeito amplia o alcance de diferentes estilos musicais, incrementando a diversidade de gêneros musicais escutados nos diferentes países do mundo. A fim de melhor entender essa diversidade, neste artigo foi construída e analisada uma rede complexa de artistas, gêneros musicais e países utilizando dados do Spotify. Como resultado, além de identificar comunidades de países com gêneros musicais semelhantes, nós mostramos como a diversidade de gêneros influencia a estrutura da rede considerada. Nós também classificamos os gêneros musicais mais comumente escutados utilizando métricas de centralidade e analisamos como ocorre a difusão de gêneros ao longo do tempo, incluindo um estudo de caso sobre música viral.*

Palavras-Chave. *Redes complexas; Ciência de redes; Popularidade de música.*

1. Introdução

A música é um tipo de arte cuja manifestação pode ser caracterizada como uma prática cultural. Seu papel está relacionado a aspectos como entretenimento e construção de identidade, memória e emoções em indivíduos [DeNora 2000]. Além disso, a música é comumente utilizada como uma forma de auto-expressão, fornecendo um meio de caracterizar e identificar, por exemplo, grupos de pessoas com interesses musicais em comum e também acontecimentos históricos [Rentfrow 2012].

Nas últimas décadas, o avanço da tecnologia tem exercido um papel importante na indústria da música, mudando a forma como ela é distribuída e consumida pelo público. Um exemplo disso consiste no surgimento das plataformas de *streaming* de música. Devido ao fato de permitirem o consumo em tempo real de músicas sem a necessidade de fazer o *download* de arquivos [Trefzger et al. 2015], essas plataformas têm recebido cada vez mais destaque. A grande adesão a esse tipo de serviço por parte dos usuários pode ser entendida como um fator que tem democratizado o acesso ao conteúdo musical. Isso pode ser observado a partir de dados do IFPI (*International Federation of the Phonographic Industry*) [IFPI 2017], que mostram que 50% da receita da indústria fonográfica em 2016 foi proveniente do uso de plataformas de *streaming* de música e *downloads*.

Basicamente, através do serviço de *streaming*, usuários escutam músicas, que por sua vez são gravadas e disponibilizadas por artistas e bandas. Através dessa estrutura, é possível ainda registrar alguns metadados que incluem: gênero dos artistas, popularidade e quantidade de *streams* de determinada música, entre outros. A coleta e o armazenamento desse tipo de informação é um aspecto importante que surge com a popularização dessas plataformas e que abre espaço para um grande conjunto de possíveis estudos. Dentre eles, destaca-se a aplicação de técnicas de ciência de redes [Barabási 2016] para a análise da estrutura das plataformas, que podem ser facilmente caracterizadas como uma rede.

Neste contexto, o presente trabalho tem como objetivo analisar a rede de gêneros musicais ao redor do mundo. Para isto, utilizamos o Spotify, uma plataforma de *streaming* de música bem consolidada e que disponibiliza os dados necessários para a construção da rede, que será composta por países, artistas e gêneros musicais. São utilizados dados sobre as músicas mais tocadas em alguns dos países onde o Spotify está disponível. Através desses dados, buscamos entender quais são os gêneros musicais mais tocados e identificar comunidades de países que compartilham interesse por gêneros musicais em comum. Neste artigo, estendemos as análises da rede, previamente apresentadas em [Mondelli et al. 2018], com um maior conjunto de dados e propomos analisar como ocorre a difusão de determinado gênero ao longo do tempo e também identificar o que caracteriza um gênero musical viral e universal. O acúmulo de conhecimento sobre um sistema é um dos objetivos da Engenharia de Software Empírica. Alinhado a esse objetivo, um melhor entendimento da demanda por conteúdo multimídia em escala global agrega conhecimento para desenvolvimento e evolução de sistemas com propósito semelhante ao sistema estudado neste artigo.

Sendo assim, este artigo está organizado como segue. A Seção 2 descreve a metodologia do trabalho, incluindo o processo de coleta de dados para a construção da rede e quais análises serão realizadas. A Seção 3 apresenta o estudo e os resultados das análises

da rede. A Seção 4 analisa a evolução e o comportamento de determinados gêneros musicais ao longo do tempo. A Seção 5 apresenta alguns trabalhos relacionados. Por fim, a Seção 6 conclui o trabalho e apresenta algumas oportunidades de trabalhos futuros.

2. Metodologia

Para a construção e estudo da rede proposta neste trabalho, seguimos uma metodologia que é composta por um fluxo de etapas. Esse fluxo que compreende desde a coleta e tratamento dos dados até a obtenção dos resultados, que serão apresentados nas seções 3 e 4. As etapas são apresentadas na figura 1 e descritas com mais detalhes nas subseções a seguir.



Figura 1. Etapas para a construção e análise da rede.

Destacamos que neste trabalho optamos por utilizar dados do Spotify, uma plataforma de *streaming* de música lançada em 2008 e que hoje reúne aproximadamente 30 milhões de músicas, 2 milhões de artistas e 140 milhões de usuários. Além de ser um serviço de *streaming* bem consolidado, o Spotify disponibiliza o acesso a parte de sua base de dados para consulta, possibilitando a construção da rede a ser analisada neste trabalho.

2.1. Coleta de dados - Spotify Charts

O Spotify Charts¹ registra, diária e semanalmente, uma lista de até 200 músicas mais escutadas nos países onde o Spotify está disponível e seus respectivos artistas. Para este trabalho, foram coletadas duas listas: a primeira referente ao dia 01/07/2017, com 9231 registros, totalizando 1326 artistas e 2438 músicas diferentes; e a segunda referente ao dia 01/01/2018, com 9219 registros e um total de 1609 artistas e 2995 músicas diferentes. Vale ressaltar que para a primeira listagem foi possível obter registros para 57 países, enquanto que para segunda foram obtidos registros para 50 países.

Pensando em uma outra possibilidade de análise, coletamos um segundo conjunto de dados, a fim de obter o registro das músicas mais escutadas ao redor do mundo para um maior período de tempo. Nesse caso, decidimos coletar os dados a partir da opção de registro semanal do Spotify Charts desde a primeira semana em que a plataforma passou a disponibilizar esse tipo de informação. Foi possível então obter um conjunto de registros para o período de 86 semanas, entre as semanas que iniciaram nos dias 23/12/2016 e 10/08/2018. Além disso, esse novo conjunto abrange um total de 62 países. No entanto, observamos que nem todos os países possuem registros para todas as 86 semanas. A fim de manter a uniformidade dos registros, e conseqüentemente das análises, optamos por excluir os 7 países que não possuíam registros para o período completo. Sendo assim, filtramos esse segundo conjunto de dados para considerar os demais 55 países.

¹<https://spotifycharts.com/>

2.2. Coleta de dados - Spotify Web API

A API do Spotify² disponibiliza metadados sobre artistas, álbuns e músicas do catálogo do Spotify. A API foi acessada a fim de recuperar os gêneros musicais de cada um dos artistas obtidos nas listagens do Spotify Charts descritas anteriormente. Vale destacar que cada artista pode possuir mais de um gênero e que, de acordo com a plataforma Every Noise [Noise 2017], o Spotify possui um registro de aproximadamente 1500 gêneros diferentes. A consulta aos metadados da API permitiu então associar cada artista ao seu respectivo gênero principal, resultando em um total de 367 diferentes gêneros para o conjunto de dados do dia 01/07/2017 e 425 gêneros para o conjunto do dia 01/01/2018. O impacto da grande quantidade de gêneros será discutido na Seção 3, onde serão feitas as análises dessas duas redes.

Para o segundo conjunto de dados, que abrange a coleta das músicas mais tocadas para um período de 86 semanas, também utilizamos a API para obter os gêneros relacionados aos artistas.

2.3. Construção da rede

A estrutura dos dados permitiu então a construção de uma rede tripartida, não direcionada, para cada uma das listagens coletadas. Nessa rede, os conjuntos de vértices podem ser classificados como: (i) países, (ii) artistas e (iii) gêneros musicais. As ligações entre as classes de vértices são apresentadas na Figura 2 e acontecem da seguinte forma:

- Países se conectam a artistas, segundo o registro obtido no Spotify Charts que relaciona cada música e seu respectivo artista aos países onde ele é mais escutado.
- Artistas se conectam aos gêneros, de acordo com os metadados obtidos através da API Web.

Com esse modelo, é possível ainda conectar cada país diretamente aos gêneros, por transitividade (linha tracejada na Figura 2). Essa conexão, que resulta na rede bipartida de país e gênero, será o objeto de estudo deste trabalho. Vale ainda ressaltar que, a construção dos conjuntos de dados e as análises apresentadas na seção a seguir, foram feitos utilizando o pacote estatístico R e estão disponíveis no GitHub³.

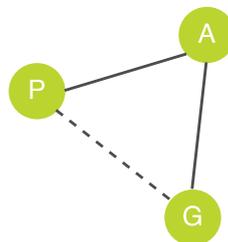


Figura 2. Exemplo da estrutura da rede de países, artistas e gêneros.

A partir desse ponto do fluxo de etapas, nós exploramos a rede de duas formas diferentes. Uma vez que a coleta de dados foi realizada e a rede foi construída, as análises que propomos a seguir podem ser feitas de forma independente. Essas análises estão indicadas na Figura 1 como Análise 1 e Análise 2 e são descritas a seguir.

²<https://developer.spotify.com/web-api>

³<https://github.com/mmondelli/network-science>

2.4. Análise 1: rede de gêneros musicais

Com a finalidade de explorar o comportamento da rede de gêneros musicais ao redor do mundo, nesta análise estudamos as duas redes construídas com os conjuntos de dados coletados dos dias 01/07/2017 e 01/01/2018. Buscamos entender o impacto da grande quantidade de gêneros musicais que existem na plataforma do Spotify e propomos uma filtragem nos dados que possibilite a detecção de comunidades de países com interesses em comum. Com essa filtragem, que também pode ser entendida como mais uma etapa da metodologia, passamos a considerar apenas o cinco gêneros mais frequentemente escutados em cada país e excluimos os gêneros que aparecem em pelo menos 90% dos países a fim de evitar as redes completamente conectadas e também de possibilitar o aparecimento de gêneros mais característicos de cada país.

2.5. Análise 2: difusão de gêneros musicais ao longo do tempo

Nesta análise, utilizamos um conjunto maior de dados que abrange um período mais extenso do registro das músicas mais tocadas ao redor do mundo. Esses dados nos permitem identificar com mais clareza questões sobre a viralidade de um determinado gênero e o que diferencia um gênero universal de um gênero viral. Em especial, utilizamos um exemplo de música que, no período de coleta de dados realizada neste trabalho, identificamos ser uma música viral. Considerando apenas se a música aparece na lista de mais tocadas em diferentes países, buscamos entender por quanto tempo o gênero acaba se tornando universal e se isso pode ou não ser influenciado por outros artistas e outras possíveis versões da música.

3. Análise da rede de gêneros musicais

Esta seção analisa a rede de gêneros musicais ao redor do mundo para cada uma das datas coletadas. Primeiramente, consideramos a rede completa de gêneros musicais e, em seguida, analisamos a rede com somente os principais gêneros musicais escutados. A fim de diferenciar os resultados das análises, nomeamos como *Rede 1* e *Rede 2* as redes que utilizam os dados dos dias 01/07/2017 e 01/01/2018, respectivamente.

3.1. Análise das redes completas de gêneros musicais

Para esta análise, consideramos o modelo de rede bipartida que relaciona os conjuntos de vértices país e gênero. A Rede 1 possui 57 países e 367 gêneros, totalizando 424 vértices e 3579 arestas. A Rede 2, por sua vez, possui 50 países e 424 gêneros, totalizando 474 vértices e 3835 arestas. Com a finalidade de entender o perfil musical de cada um dos países e analisar as características das duas redes, foi realizada a projeção da rede de países conectados por gêneros em comum. Essa projeção resultou em uma rede completamente conectada, onde cada país está conectado aos demais, como mostra a Figura 3.

Analisando os dados, observa-se que isso ocorre devido a existência de pelo menos um gênero que ocorre em todos os países. Esse tipo de gênero pode ser classificado então como sendo um gênero universal, que é o caso do gênero *pop*, por exemplo. Além da existência de gêneros universais, outro aspecto pode ser observado: o fator viral de uma determinada música ou artista pode fazer com que um gênero também ocorra em todos os países, mesmo que pontualmente. Um exemplo disso consiste no gênero *latin*, que

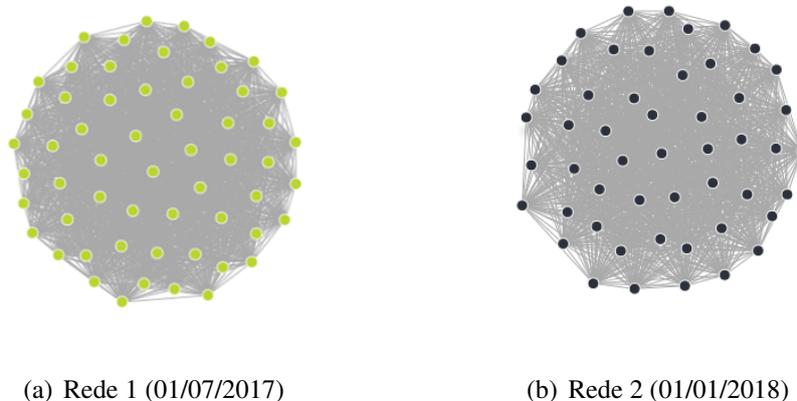


Figura 3. Projeção de países conectados por gênero em comum.

aparece em todos os países da Rede 1 devido à popularidade da música “Despacito” do cantor Luis Fonsi, identificada como a música mais escutada ao redor do mundo. Nesse caso, é possível atribuir parte da sua popularidade ao aumento do uso de plataformas de *streaming* de música e vídeo⁴.

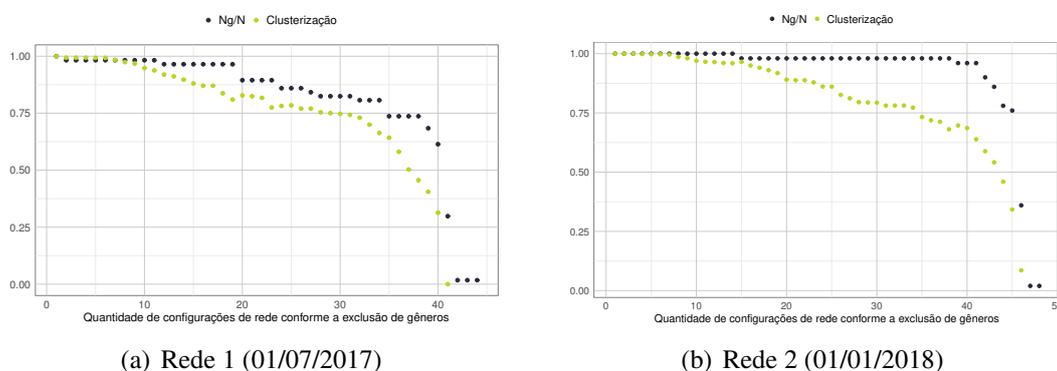


Figura 4. Comportamento da rede com a exclusão de gêneros.

Considerar a estrutura original dos conjuntos de dados deste trabalho impossibilita o estudo das nuances acerca do perfil musical de cada um dos países, devido a existência de gêneros universais e músicas virais. Foi então realizada uma análise das Redes 1 e 2, a fim de entender seus comportamentos conforme uma determinada quantidade de gêneros fosse excluída. A exclusão foi feita iterativamente, de forma com que no primeiro passo fossem excluídos todos os gêneros que aparecem em todos os n países, onde n é o número total de países de cada rede. Depois foram excluídos os gêneros que aparecem em pelo menos $n - 1$ países e assim por diante. Esse processo foi repetido até que não existissem mais gêneros a serem excluídos, totalizando 44 passos (ou redes diferentes) para a Rede 1 e 48 passos para a Rede 2. Foram contabilizados a fração de vértices do componente

⁴<http://www.billboard.com/articles/columns/latin/7873798/luis-fonsi-daddy-yankee-despacito-streaming-popularity>

gigante (N_g) em relação ao número total de vértices de cada rede (N) e a clusterização em cada caso. O resultado é apresentado na Figura 4. É possível perceber que, em ambos os casos, a clusterização tende a permanecer alta. O componente gigante abrange a maioria dos nós da rede por muitos passos, diminuindo de forma mais significativa apenas quando restam poucos gêneros a serem excluídos. Sendo assim, escolher um limite de gêneros a ser considerado nesse caso para prosseguir com a caracterização da rede não é trivial. Neste trabalho, optou-se então por analisar apenas os gêneros mais frequentes em cada um dos países, conforme descreve a subseção a seguir.

3.2. Análise da rede para os 5 gêneros musicais mais escutados

Como observado na análise das redes com a exclusão de gêneros, existe uma sensibilidade no que diz respeito ao estado das redes que pode ser considerado para a continuação do estudo. Sendo assim, neste trabalho optou-se por fazer uma filtragem nos dados, como descrita na Seção 2, de forma a considerar apenas os cinco gêneros mais frequentemente escutados em cada país. Vale ressaltar também que, para a construção destas redes, foram excluídos gêneros que aparecem em pelo menos 90% dos países.

A rede bipartida de países e gêneros, no caso da Rede 1, possui 157 vértices e 259 arestas. Para o caso da Rede 2, possui 155 vértices e 250 arestas. As redes podem ser analisadas a partir de suas projeções ponderadas. A Tabela 1 mostra as características das projeções, onde: (i) *Países* diz respeito à projeção de países conectados por gêneros em comum; e (ii) *Gêneros* é a projeção de gêneros conectados por países em comum. Em comparação ao que é esperado para redes aleatórias equivalentes às duas projeções das Redes 1 e 2, observa-se que ambas possuem distância e diâmetro baixos e alta clusterização. Além disso, é possível observar que as redes são bem conectadas, com grande parte dos vértices pertencendo ao maior componente conexo.

Tabela 1. Características das projeções da rede bipartida de gêneros e países (top 5).

Características	Rede 1		Rede 2	
	Países	Gêneros	Países	Gêneros
Vértices	56	101	50	105
Arestas	459	310	357	327
Grau máximo	36	26	33	40
Grau mínimo	0	3	0	4
Grau médio	16.3	6.1	14.3	6.2
Distancia média	1.76	2.6	2.06	2.8
Diâmetro	5	7	9	9
Clusterização	0.79	0.43	0.8	0.39
Densidade	0.29	0.06	0.29	0.005
Componentes	6	6	3	3
Tamanho maior componente	91%	75%	96%	90%

3.2.1. Ranqueamento dos gêneros musicais

A partir da projeção de gêneros conectados por países em comum, foi possível obter o ranqueamento dos gêneros mais centrais para as Redes 1 e 2. As Tabelas 2 e 3 apresentam uma listagem dos primeiros 5 gêneros classificados de acordo com três diferentes métricas de centralidade descritas a seguir:

- **Grau:** é considerada a medida mais simples de centralidade em redes e está relacionada à quantidade de conexões que um determinado vértice possui. Sendo assim, quanto mais conexões, maior a importância do vértice na rede. Para redes direcionadas, a centralidade de grau leva em consideração a direção da conexão.
- **Closeness:** calcula o comprimento médio dos caminhos mais curtos de um vértice para cada um dos demais vértices que compõem a rede. Quanto maior a centralidade de um determinado vértice, menor é a sua distância para os demais.
- **Betweenness:** estabelece a importância de um vértice com base na quantidade de caminhos mínimos pelos quais ele faz parte. Em outras palavras, esse tipo de centralidade quantifica o número de vezes em que um vértice atua como ponte no caminho mais curto entre dois outros vértices.

A respeito dos resultados de ranqueamento obtidos, vale ressaltar que a frequência com que esses gêneros aparecem em cada país não está sendo levada em consideração, mas sim se eles aparecem pelo menos uma vez, de acordo com a filtragem proposta nesta seção. O gênero *southern hip hop*, por exemplo, é mais frequente nos Estados Unidos e Canadá, enquanto que o *reggaetown* é frequente nos países da América Latina em geral. Em um trabalho futuro, uma oportunidade estaria em considerar a frequência dos gêneros nas análises.

Por se tratar de uma rede bipartida, a interpretação dos ranqueamentos apresentados nas Tabelas 2 e 3 deve levar em consideração o fato de que os gêneros estão conectados se possuem pelo menos um país em comum. Através dos dados da Rede 1 foi possível identificar que o gênero *southern hip hop* está presente em apenas 14 países, enquanto que o *reggaeton* aparece em 22 países. No entanto, comparando os dois casos, os países que escutam *southern hip hop* são mais distintos entre si, fazendo com que ele se conecte a mais gêneros e por isso possua maior centralidade de grau. Esse mesmo aspecto foi observado através de uma varredura nos dados da Rede 2, onde existem gêneros como *vegas indie* e *progressive house* que estão presentes em mais países do que os listados no ranqueamento, mas que não possuem alta centralidade de grau. Outro fator que pode ser observado é a similaridade dos resultados das métricas *closeness* e *betweenness*, principalmente para a Rede 1. Esses resultados para *closeness* e *betweenness* sugerem que os gêneros listados atuam tanto aproximando os países com estilos musicais em comum, quanto conectando países que possuem maior diversidade de gêneros musicais, respectivamente.

Além disso, é possível observar a classificação de gêneros como *west coast rap*, *latin hip hop*, *trap music* e *trap latino* dentre os gêneros mais centrais, mesmo com a exclusão dos gêneros *rap*, *hip hop* e *latin*. Isso mostra que, de alguma forma, esses últimos três gêneros ainda estão presentes através de suas variações, indicando sua importância

na rede. Isso também ocorre com o *neo mellow* que, de acordo com mapa construído pelo Every Noise [Noise 2017], é um gênero muito similar ao *pop*, que também não está presente em nenhuma das redes.

Tabela 2. Ranqueamento dos gêneros musicais para a Rede 1.

	Grau	Closeness	Betweenness
1	southern hip hop	reggaeton	reggaeton
2	reggaeton	neo mellow	neo mellow
3	neo mellow	southern hip hop	southern hip hop
4	west coast rap	west coast rap	deep tropical house
5	tropical	deep tropical house	west coast rap

Tabela 3. Ranqueamento dos gêneros musicais para a Rede 2.

	Grau	Closeness	Betweenness
1	trap music	southern hip hop	trap music
2	big room	trap music	southern hip hop
3	tropical	rock	viral pop
4	southern hip hop	deep funk carioca	tropical
5	viral pop	viral pop	big room

3.2.2. Países com interesses musicais em comum

A identificação de comunidades em ciência de redes pode ser entendida como a busca por grupos de nós que possuem maior probabilidade de se conectarem entre si do que a nós de outras comunidades [Barabási 2016]. Sendo assim, comunidades reúnem um conjunto de nós que estão mais conectados internamente do que com outros nós em uma rede aleatória equivalente. Partindo dessa definição e da projeção de países conectados por gêneros, este trabalho buscou identificar os grupos de países que possuem interesses em comum. Para isso, foi aplicado o método de Louvain [Blondel et al. 2008], que possui uma abordagem baseada na otimização da modularidade e é considerado o estado da arte para detecção de comunidades em redes. Basicamente, em um primeiro momento, o método atribui cada vértice a uma comunidade. O método segue de forma iterativa, reatribuindo os vértices às comunidades de forma com que eles se movam para as comunidades onde contribuem mais para a modularidade. Quando nenhum vértice pode ser mais reatribuído, o processo continua, mas considerando as comunidades como sendo os vértices. O processo é finalizado quando a modularidade global máxima é encontrada.

Como resultado, na Rede 1 foram encontrados 8 comunidades de países. Dessas comunidades, 5 são comunidades individuais de países que não estavam conectados ao maior componente da rede, sendo eles: Japão, Finlândia, Noruega, Turquia e Suécia. Com a filtragem dos 5 gêneros mais frequentes proposta na análise desta seção, esses países passaram a não compartilhar nenhum gênero com os demais. No caso do Japão por exemplo, os gêneros mais comuns são o *anime*, juntamente com variações regionais

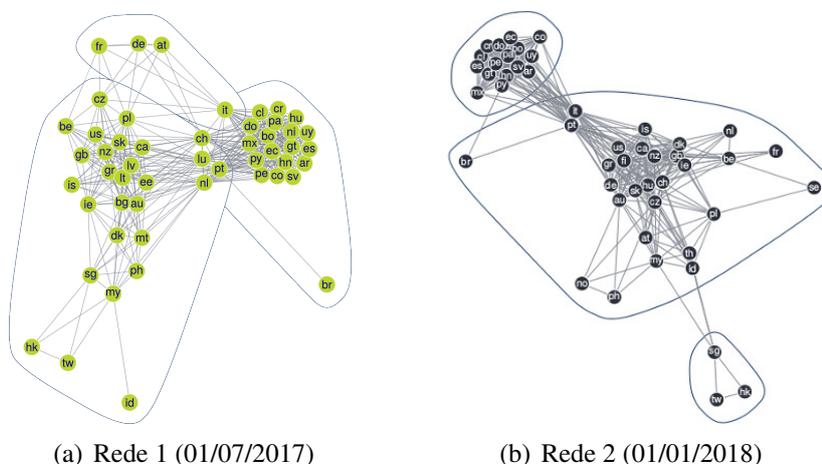


Figura 5. Comunidades de países identificada pelo método de Louvain.

do *pop*, *rap*, *r&b* e *rock*. As outras 3 comunidades são apresentadas na Figura 5(a) e descritas a seguir:

1. Argentina, Bolívia, Chile, Colômbia, Costa Rica, República Dominicana, Equador, Espanha, Guatemala, Honduras, México, Nicarágua, Panamá, Perú, Paraguai, São Salvador e Uruguai, Brasil, Portugal, Hungria e Suíça.
2. Áustria, Alemanha, Itália e França.
3. Estados Unidos, Reino Unido, Austrália, Bélgica, Bulgária, Canadá, República Checa, Dinamarca, Estônia, Grécia, Holanda, Hong Kong, Indonésia, Irlanda, Islândia, Lituânia, Letônia, Luxemburgo, Malta, Malásia, Nova Zelândia, Filipinas, Polônia, Singapura, Eslováquia e Taiwan.

É interessante notar que, na comunidade 1 foram claramente agrupados a maioria dos países de língua espanhola. A comunidade 2 agrupou alguns países da Europa. A comunidade 3 por sua vez é a maior e mais diversificada, possuindo países da Europa, Ásia, América do Norte e Oceania.

Para a Rede 2 foi identificado um total de 5 comunidades, onde 2 são comunidades individuais dos países Japão e Turquia. No caso da Rede 1, esses dois países também não fazem parte de nenhuma outra comunidade, sugerindo que de fato eles possuem predileção por estilos musicais próprios. As demais comunidades, apresentadas na Figura 5(b), são listadas a seguir:

1. Argentina, Bolívia, Chile, Colômbia, Costa Rica, República Dominicana, Equador, Espanha, Guatemala, Honduras, México, Panamá, Peru, Paraguai, São Salvador e Uruguai.
2. Hong Kong, Singapura e Taiwan.
3. Estados Unidos, Reino Unido, Áustria, Austrália, Bélgica, Brasil, Canadá, Suíça, República Checa, Alemanha, Dinamarca, Finlândia, França, Grécia, Hungria, Indonésia, Irlanda, Islândia, Itália, Malásia, Holanda, Noruega, Nova Zelândia, Filipinas, Polônia, Portugal, Suécia, Eslováquia e Tailândia.

Observa-se que, diferentemente do resultado da Rede 1, a comunidade 1 não agrupa nenhum outro país que não seja de língua espanhola. A comunidade 2 passou a agrupar

países da Ásia e a comunidade 3 continua sendo a maior e mais diversificada.

Apesar dos agrupamentos, em ambos os casos percebe-se ainda que muitos países de uma comunidade se conectam a países de outras comunidades, podendo indicar que de fato eles são parecidos musicalmente ou ser consequência da grande quantidade de gêneros diferentes que existem mesmo com os filtros aplicados. No geral, observamos que o comportamento dos agrupamentos nas Redes 1 e 2 mantém um padrão. No entanto, as diferenças pontuais podem acontecer devido a filtragem de gêneros que foi proposta nesta seção e também ser consequência das músicas que são listadas como populares em cada um dos períodos. A identificação de comunidades nesse tipo de rede pode também sugerir a existência de zonas de influência de determinados gêneros em grupos de países, podendo ser objeto de estudo em análises mais detalhadas no futuro.

Estendemos as análises apresentadas nesta seção para um conjunto maior de dados, que abrange os meses intermediários entre os meses da Rede 1 (julho) e Rede 2 (janeiro), totalizando então 5 novas redes. No geral, foi possível perceber que a estrutura das novas redes, após a aplicação dos filtros descritos no início da seção, apresentam características similares às apresentadas na Tabela 1. Em relação ao ranqueamento de gêneros musicais, apesar da grande variação dos resultados, observamos a presença do gênero *neo mellow*, uma variação do *pop*, em grande parte das classificações. As configurações de comunidades também apresentam variações. No entanto, observamos que o agrupamento de países de língua espanhola permanece ao longo dos meses, indicando que existe uma uniformidade de preferências musicais entre esta comunidade.

3.2.3. Quantos gêneros de um determinado país são de fato dele?

A plataforma Every Noise desenvolveu um mecanismo onde, além de identificar todos os gêneros musicais existentes a partir de características das músicas (incluindo aspectos sobre acústica, energia, entre outros), é possível atribuir os gêneros ao seu país considerado origem. Essa informação está também disponível através de seu site. Considerando então os 5 gêneros mais frequentes em cada um dos países, é possível determinar qual é a proporção desses gêneros que são de fato locais.

Essa análise foi realizada para as Redes 1 e 2 mas, devido à similaridade dos resultados, optou-se por apresentar apenas um deles. Sendo assim, para cada país presente na Rede 1, os gêneros da rede foram comparados com a relação de gêneros e países da plataforma Every Noise. A Figura 6 mostra o resultado dessa comparação, onde 100% indica que os 5 gêneros de um país são de origem dele. Dos 57 países da Rede 1, apenas 12 possuem pelo menos 80% de gêneros considerados originais. Ou seja, pode-se dizer que, eliminando os gêneros universais existe pouca influência de outros países no estilo musical desse conjunto de 12 países. Por outro lado, um grande conjunto de países são completamente influenciados ou importam gêneros musicais dos demais. O resultado desse tipo de análise pode também estar relacionado à produção musical própria de um determinado país. Em outras palavras, países que possuem uma indústria musical bem estabelecida podem estar produzindo um conteúdo mais voltado às suas questões culturais e costumes. Um exemplo disso é o Brasil, onde o forró, o *funk* carioca e o sertanejo e

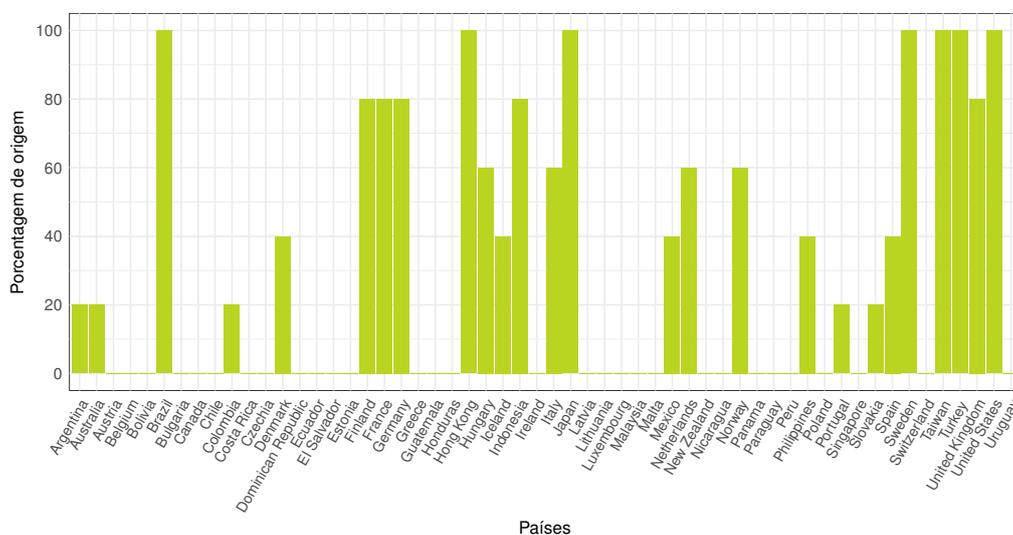


Figura 6. Proporção de gêneros que são originados em cada país.

suas variações foram classificados como os gêneros mais escutados. Esse tipo de análise permite que estudos futuros busquem entender como ocorre a influência e o fluxo de estilos musicais entre países, identificar quais são os países mais influentes e explorar a estrutura da produção musical em cada país, por exemplo.

4. Análise da difusão de gêneros musicais ao longo do tempo

Na Seção 3.1 identificamos que a existência de gêneros universais e virais podem levar a rede de gêneros e países a estar completamente conectada. Assim, esta seção se propõe a analisar esse aspecto de forma mais detalhada.

A partir do segundo conjunto de dados, que compreende o registro de músicas mais tocadas num período de 86 semanas, buscamos uma forma de demonstrar o que caracteriza e diferencia gêneros universais de gêneros virais. Uma possibilidade está em contabilizar o número de países em que um determinado gênero ocorre, ao longo das semanas. Na Figura 7 apresentamos um exemplo para os gêneros pop e latino. Podemos observar que, de fato, o gênero latino não é universal, pois ele não ocorre em todos os países em todo o período, como ocorre com o pop. Vale ressaltar, contudo, que consideramos como ocorrência do gênero se ele aparece pelo menos uma vez na listagem das 200 músicas de cada país por semana.

Neste trabalho, utilizamos o gênero latino para exemplificar o aspecto de viralidade pois, para o período coletado, sabemos que esse gênero ganhou um grande destaque devido à música *Despacito*, de Luis Fonsi, citada também na Seção 3.1. Segundo fontes de notícia [Veja 2017, Estadão 2017], o videoclipe da música foi o mais visto no YouTube, passando de 5 bilhões de visualizações em março de 2018. Além de ter sido indicada a várias premiações, a música foi posicionada entre as dez melhores músicas em 2017 pelas revistas *Time*, *Billboard* e *Rolling Stone*. Assim, complementando a análise apresentada na Figura 7 e com foco apenas no gênero latino, mostramos na Figura 8 o comportamento

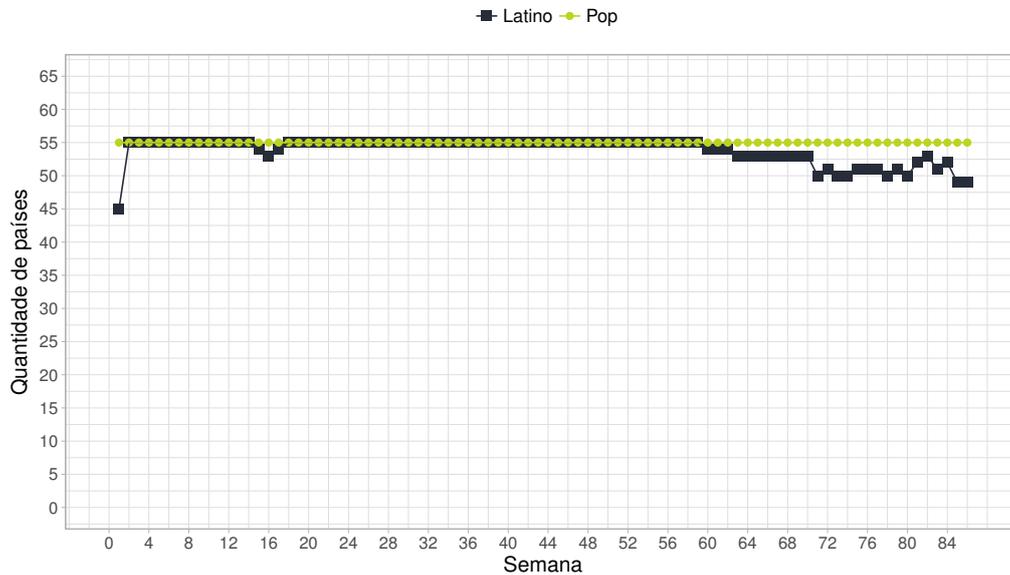


Figura 7. Quantidade de países em que os gêneros pop e latino ocorrem por semana.

do gênero ao longo das semanas, contabilizando o número de países em que o gênero ocorre com ou sem a música Despacito. Podemos observar que, em certos períodos (entre as semanas 8 e 25) a música é responsável por tornar o gênero universal.



Figura 8. Quantidade de países em que o gênero latino ocorre por semana, considerando ou não a música Despacito.

Analisando apenas a música Despacito, podemos observar ainda sua evolução ao longo do tempo, como mostra a Figura 9. A música foi originalmente lançada em 12/01/2017, e sua variação, um *remix* com participação do cantor Justin Bieber, lançada

em 21/04/2017. A partir dessa análise temos que:

- Despacito original: surge na semana 4, se torna universal na semana 7 até 14 (7 semanas), sofre uma diminuição, e volta a ser universal na semana 22 até 38 (16 semanas). No total, a música permaneceu universal durante 23 semanas.
- Variação (*remix*): surge na semana 16, passa a ser universal na semana 18 e permanece até a semana 43, totalizando 25 semanas.

É interessante perceber que nas semanas 15 e 16 a música original apresenta uma pequena queda, mas com o surgimento do *remix* volta a ser universal nas semanas seguintes. Outro aspecto importante a ser considerado em relação ao *remix* é que, nesse caso, ele contou com a participação de um cantor bastante popular do gênero pop. Sabemos, da análise apresentada na Figura 7, que o pop é um gênero universal e esse fator influencia o aparecimento dele em todos os países. O gráfico mostra ainda que a versão *remix* da música permanece universal mesmo depois da música original ter deixado de estar presente em todos os países. Assim, o lançamento e aumento da popularidade do *remix* pode ser entendido como um fator que impulsionou a popularidade e estendeu o efeito de viralidade da música e, conseqüentemente do gênero latino, ou ainda que sustentou esse efeito por um período maior de tempo.

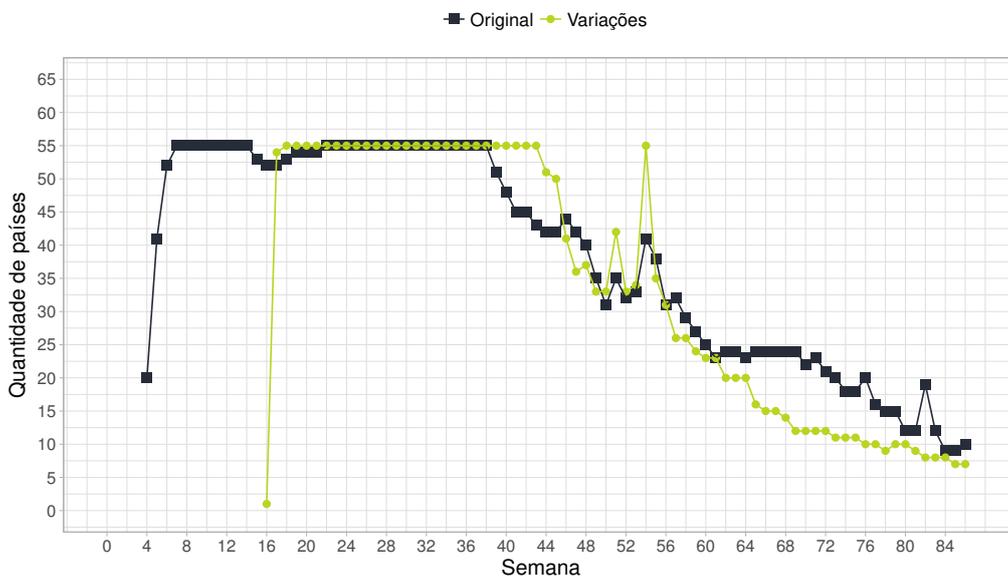


Figura 9. Quantidade de países que a música Despacito e sua variação alcançaram ao longo do tempo.

Comparando as Figuras 8 e 9, podemos observar que mesmo sem considerar a música Despacito, o gênero latino está presente em todos os países entre as semanas 2 a 6. O mesmo ocorre entre as semanas 44 e 58, onde sabemos que tanto Despacito quanto o *remix* já deixaram de ser universais. Esse aspecto é um indicador de que outras músicas do gênero surgiram nesse período e se tornaram também virais, alcançando todos os países. De acordo com os dados coletados, essas músicas estão descritas na Figura 10. Podemos observar que entre o intervalo de semanas 51 e 53 não houve outras músicas latinas virais

e, conforme mostra a Figura 8, o gênero latino continuou viral nesse mesmo período. Por isso, ressaltamos que o gênero pode também estar presente em todos os países através de músicas que não são virais.

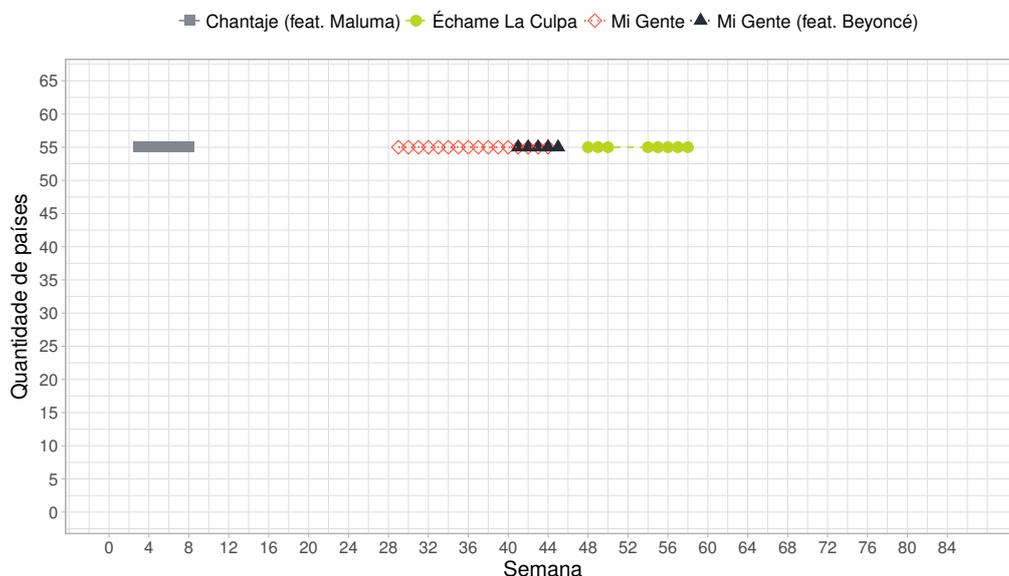


Figura 10. Outras músicas latinas virais no período de consulta dos dados.

Através dos dados foi possível perceber também que, nos casos apresentados na Figura 10, os artistas também podem ser classificados como sendo do gênero pop, como por exemplo a cantora Shakira que possui tanto o gênero pop quanto o latino. Podem existir ainda casos em que um artista do gênero latino possua uma música com participação de um artista de gênero pop, como ocorreu com o *remix* de Despacito e também é o caso da música Mi Gente de J Balvin em parceria com a cantora Beyoncé. Desta forma, esses dois últimos fatores ilustram outras maneiras em que gêneros não universais, como no caso do gênero latino, podem se tornar virais ou temporariamente universais.

5. Trabalhos relacionados

No contexto de sistemas multimídia, os avanços nas tecnologias de rede de computadores e processamento de dados permitiram o surgimento de plataformas de *streaming* de conteúdo através da internet. O surgimento dessas plataformas aliado ao aumento da utilização de dispositivos móveis por parte dos usuários, permitem acesso mais rápido e fácil ao conteúdo multimídia. Atualmente, as plataformas de *streaming* que têm se destacado no mercado incluem Netflix e Amazon Prime, para *streaming* de vídeo, e Apple Music e Spotify, para *streaming* de música. Na literatura, são comuns trabalhos focadas na análise de desempenho dessas plataformas ou na análise do impacto do tráfego de *streaming* na estrutura da rede. De forma mais específica, é possível encontrar alguns trabalhos que buscam estudar e analisar aspectos sobre a utilização dessas plataformas. No entanto, encontramos poucas contribuições com um foco maior na análise da estrutura de redes que podem ser construídas a partir de plataformas de *streaming* de música mais

especificamente, tal como realizado neste trabalho. Os estudos que mais se aproximam da abordagem proposta neste trabalho são apresentados a seguir.

O trabalho proposto por [Lee and Cunningham 2012] estuda o fluxo de música entre diferentes regiões do mundo, utilizando dados da plataforma de *streaming* de música Last.fm. O estudo busca entender quais são as preferências musicais de cada região e determinar os agrupamentos entre elas. São estabelecidas relações do tipo líder-seguidor entre os nós, permitindo entender qual é a dinâmica da preferência musical na rede. No entanto, como visto neste trabalho, a existência da grande quantidade de gêneros é um fator que influencia a estrutura de redes que visam estudar preferências musicais entre países. Em [Lee and Cunningham 2012] não fica claro qual foi a quantidade de gêneros considerada para o estudo e se houve ou não algum tipo de filtragem.

Em [Bryan and Wang 2011], a partir de dados da plataforma WhoSampled, são classificados os gêneros e artistas mais influentes no mercado musical de *sampling*. *Sampling* diz respeito à seleção de uma amostra de gravação de som a fim de reutilizá-la em uma outra gravação diferente. Sendo assim, um registro no conjunto de dados consiste em um artista que teve sua música amostrada e o artista que utilizou a amostra, incluindo metadados como o gênero. Esse tipo de dado possibilitou a construção de três redes direcionadas de músicas, artistas e gêneros. Em comparação com o presente trabalho, a abordagem em questão não aborda a grande diversidade de gêneros e difere no que diz respeito à estrutura da rede e ao tipo de dado que foi analisado.

A diversidade de gêneros é abordada em [Lambiotte and Ausloos 2006] com o foco no comportamento dos usuários em geral, e não nos países, como foi feito neste trabalho. São utilizados dados do Last.fm a fim de identificar padrões coletivos a partir das músicas escutadas. O trabalho identificou um total de 142 gêneros distintos, divididos em grupos musicais. A partir da correlação entre os gêneros, obteve-se um mapa onde foi possível observar que determinados gêneros musicais tendem a ser escutados pelos mesmos usuários. Em [Lambiotte and Ausloos 2005] a estrutura dessa mesma rede é explorada de forma aleatória a fim de mapear a estrutura interna da rede e as correlações em uma série temporal. A análise estatística dessa série permitiu identificar comportamentos não-triviais de usuários sugerindo, por exemplo, que eles se tornam mais ecléticos musicalmente.

A busca por comunidades em redes de música é também estudada em [Smith 2006] e [Gleiser and Danon 2003]. No entanto, ambos os trabalhos levam em consideração gêneros musicais específicos. Em [Smith 2006] é analisada a estrutura da rede de colaboração entre cantores de *rap* e em [Gleiser and Danon 2003] é analisada a rede de cantores de *jazz*. Neste último foi possível identificar a segregação racial entre as comunidades de artistas encontradas.

Em [Vlegels and Lievens 2017] é construída e analisada uma rede para identificar padrões de interesse musical. No entanto, em [Vlegels and Lievens 2017] não são utilizados dados de plataformas de *streaming*, mas sim de um levantamento feito com participantes de uma pesquisa na Bélgica. São levados em consideração outros atributos mais focados nos participantes, que compreendem, por exemplo, dados socioeconômicos. Além disso, o trabalho está limitado a uma quantidade pequena de apenas 17 gêneros musicais.

Também voltado para a identificação de preferências musicais, em [Pereira et al. 2018] são utilizadas técnicas de análise de redes sociais temporais para estudar a evolução do gosto musical de usuários da plataforma *This Is My Jam*. O trabalho, no entanto, foca na relação de similaridade entre os usuários e entre os artistas, não levando em consideração questões sobre o gênero das músicas.

Em contraste, o presente trabalho apresentou de forma mais ampla uma visão da rede de gêneros musicais escutados ao redor do mundo a partir dos dados do Spotify. O Spotify se destaca pela sua popularidade e pelo seu esforço em aprimorar a experiência do usuário no que diz respeito ao consumo de conteúdo musical. Em comparação com as demais abordagens, ressaltamos que o Spotify classifica os gêneros musicais através de algoritmos que utilizam informações específicas das músicas, tais como força da batida e energia. Sendo assim, a estrutura da rede analisada reflete a grande quantidade de novos gêneros descobertos pela plataforma e não exclui as características intrínsecas a cada um deles.

6. Conclusões e trabalhos futuros

Neste trabalho, foi construída uma rede de países e gêneros com base no registro de músicas mais escutadas em cada país disponibilizado pelo Spotify. O modelo de rede foi estudado a partir diferentes conjuntos de dados, coletados através da plataforma para períodos diferentes. Dessa forma, demonstramos que a metodologia utilizada para as análises não se restringe à uma data específica. Ressaltamos, no entanto, que o trabalho possui algumas limitações. Uma delas está relacionada à disponibilidade da listagem de músicas mais escutadas ao redor do mundo, registrada pelo Spotify Charts. Identificamos que a plataforma atualiza o período em que os dados podem ser consultados e isso pode restringir a identificação de outras músicas virais, por exemplo. Para o presente trabalho, não foi possível coletar dados no período anterior à dezembro de 2016. Outra limitação está no registro de gêneros dos artistas, que não é feito de maneira uniforme. Observamos que artistas com menos popularidade podem não ter gêneros associados a eles. Esse aspecto não impactou de forma significativa este trabalho, mas entendemos que uma nova análise com foco nesses artistas teria restrições ou demandaria algum tipo de tratamento especial de dados.

A análise da projeção de países conectados por gêneros em comum permitiu identificar o impacto que a quantidade de gêneros exerce sobre a estrutura da rede, com base nos conjuntos de dados utilizados. Gêneros considerados universais e virais foram responsáveis por conectar todos os países entre si. Uma vez que o trabalho teve como objetivo entender o estilo musical dos países, foi feita uma filtragem que buscou excluir os gêneros universais e, a partir disso, considerar apenas os 5 gêneros mais frequentes em cada país. Com essa nova configuração de rede, foi possível classificar os gêneros não universais segundo métricas de centralidade de redes. A classificação mostra que, mesmo com a exclusão dos gêneros universais, esses mesmos gêneros continuam presentes através de suas variações (ou subgêneros), demonstrando sua importância no cenário musical como um todo. Além disso, com a aplicação do método de Louvain foram identificadas comunidades de países que compartilham interesses musicais em comum. Para os dois conjuntos de dados utilizados, foi interessante observar o surgimento de uma co-

munidade que agrupou países de língua espanhola. Esse tipo de comunidade destaca que, apesar da língua falada nesses países não ter sido um atributo levado em consideração no conjunto de dados, esses países tendem a ter estilos musicais parecidos. Trabalhos futuros poderiam incluir as características específicas dos gêneros, como ocorre em sistemas de recomendação e classificação [Shakya et al. 2017], a fim de se obter uma análise mais detalhada sobre as relações musicais entre países.

Uma das análises do trabalho identificou qual a proporção de gêneros escutada por um país é de fato dele. Essa análise mostra que para um grande conjunto de países, os gêneros são provenientes de outros. Um trabalho futuro pode analisar como ocorre esse fluxo de gêneros entre países. Também como trabalho futuro, observa-se a possibilidade de analisar a rede com foco nos artistas. Essa perspectiva permitirá, por exemplo, identificar quais são os artistas responsáveis pelo surgimento de determinados gêneros nos países e classificar os artistas mais influentes tanto global quanto regionalmente.

Por fim, neste trabalho apresentamos um conjunto de análises que possibilitaram observar como ocorre a difusão de determinado gênero conforme sua popularidade aumenta e qual é a duração desse fator viral. Estudamos o gênero latino como exemplo pois, durante o desenvolvimento do trabalho e de acordo com os dados coletados, foi um gênero que se destacou devido a popularidade da música Despacito. Observamos que o fator viral, nesse caso, foi influenciado e sustentado por um período maior também pela popularidade de uma segunda versão da música, lançada algumas semanas depois da música original. Além disso, identificamos que a popularidade de um determinado gênero pode estar relacionada a outros fatores, como a participação de cantores de gêneros universais nas músicas. A dinâmica de demanda de conteúdo em plataformas de *streaming* traz, portanto, desafios que impactam seu desempenho e eficiência na entrega do serviço proposto. Portanto, um melhor entendimento das diferentes perspectivas sob as quais essa demanda se apresenta tem o potencial de contribuir para o desenvolvimento e evolução de sistemas com objetivo similar. Uma possibilidade de trabalho futuro estaria em integrar dados de outras plataformas, como Twitter e Youtube, a fim de identificar novas relações e prever a duração desse fator viral.

Agradecimentos

Agradecemos o suporte da CAPES, CNPq, FAPERJ, FAPESP e do INCT em Ciência de Dados (INCT-CiD).

Referências

- Barabási, A.-L. (2016). *Network science*. Cambridge University Press.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.
- Bryan, N. J. and Wang, G. (2011). Musical influence network analysis and rank of sample-based music. In *ISMIR*, pages 329–334.
- DeNora, T. (2000). *Music in everyday life*. Cambridge University Press.

- Estadão (2017). A história real por trás de despacito, o hit do momento. <https://cultura.estadao.com.br/noticias/musica,a-historia-real-por-tras-de-despacito-o-hit-do-momento,70001836771>. Acessado em: 14/09/2018.
- Gleiser, P. M. and Danon, L. (2003). Community structure in jazz. *Advances in complex systems*, 6(04):565–573.
- IFPI (2017). IFPI facts and stats. <http://www.ifpi.org/facts-and-stats.php>. Acessado em: 01/09/2017.
- Lambiotte, R. and Ausloos, M. (2005). Uncovering collective listening habits and music genres in bipartite networks. *Physical Review E*, 72(6):066107.
- Lambiotte, R. and Ausloos, M. (2006). On the genre-fication of music: a percolation approach. *The European Physical Journal B-Condensed Matter and Complex Systems*, 50(1):183–188.
- Lee, C. and Cunningham, P. (2012). The geographic flow of music. In *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*, pages 691–695. IEEE Computer Society.
- Mondelli, M. L. B., Gadelha Jr, L. M., and Ziviani, A. (2018). O que os países escutam: Analisando a rede de gêneros musicais ao redor do mundo. In *7º Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2018)*, volume 7, pages 148–159. SBC.
- Noise, E. (2017). Every Noise at Once. <http://everynoise.com/engenremap.html>. Acessado em: 01/09/2017.
- Pereira, F. S. F., Linhares, C. D. G., Ponciano, J. R., Gama, J., de Amo, S., and Oliveira, G. M. B. (2018). That’s my jam! uma análise temporal sobre a evolução das preferências dos usuários em uma rede social de músicas. In *7º Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2018)*, volume 7, pages 160–171. SBC.
- Rentfrow, P. J. (2012). The role of music in everyday life: Current directions in the social psychology of music. *Social and personality psychology compass*, 6(5):402–416.
- Shakya, A., Gurung, B., Thapa, M. S., Rai, M., and Joshi, B. (2017). Music classification based on genre and mood. In *International Conference on Computational Intelligence, Communications, and Business Analytics*, pages 168–183. Springer.
- Smith, R. D. (2006). The network of collaboration among rappers and its community structure. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(02):P02006.
- Trefzger, T., Rose, M., Baccarella, C., and Voigt, K.-I. (2015). Streaming killed the download star! how the business model of streaming services revolutionizes music distribution. *Journal of Organizational Advancement (Online), Strategie and Institutional Studies*, 7(1):29–39.
- Veja (2017). Despacito é a música mais tocada na história do streaming. <https://veja.abril.com.br/entretenimento/>

despacito-e-a-musica-mais-tocada-na-historia-do-streaming.
Acessado em: 14/09/2018.

Vlegels, J. and Lievens, J. (2017). Music classification, genres, and taste patterns: A ground-up network analysis on the clustering of artist preferences. *Poetics*, 60:76–89.