

Difusão do comportamento de engajamento musical na plataforma Last.fm

Matheus Henrique B. dos Santos¹, Rian Wagner Costa¹, Henrique M. C. Hott¹,
Carolina Ribeiro Xavier¹, Vinícius da Fonseca Vieira¹

¹Departamento de Ciência da Computação – Universidade Federal de São João del-Rei
São João del-Rei, MG, Brasil

vinicius@ufsj.edu.br

Abstract. *The information diffusion processes in online social networks are directly affected by how individuals interact with each other, as well as how they adopt behaviors on these platforms. Understanding how individuals act and propagate their actions in social networks is of fundamental importance in various contexts, such as politics, economics, and culture. This study aims to characterize and model user engagement with music by considering their social interactions, which define a network, and their musical habits, which define their profile. Applying the methodology to the context of Linkin Park's musics on the Last.fm platform helps clarify the roles of user profiles and their interrelations in the diffusion of musical engagement behavior.*

Resumo. *Processos de difusão de informação em redes sociais online são diretamente afetados pela forma como indivíduos se relacionam uns com os outros, além da forma como adotam comportamentos nessas plataformas. A compreensão da forma como indivíduos agem e propagam suas ações em redes sociais é de fundamental importância em diversos contextos, como política, economia e cultura. Este trabalho tem como objetivo caracterizar e modelar o engajamento de usuários a músicas considerando suas interações sociais, que definem uma rede, e seus hábitos musicais, que definem seu perfil. A aplicação da metodologia ao contexto de músicas da banda Linkin Park na plataforma Last.fm ajudam a esclarecer quais os papéis dos perfis dos usuários e das suas inter-relações na difusão do comportamento de engajamento musical.*

1. Introdução

Com a crescente transformação digital da nossa sociedade, tem havido um grande aumento no engajamento de usuários em plataformas de redes sociais *online*, contribuindo para o espalhamento de informações através desses ambientes [Al-Garadi et al. 2018, Badawy et al. 2018, Tasente 2020, Allcott and Gentzkow 2017, Ferreira et al. 2021]. Processos de difusão de informação são diretamente afetados pela forma como indivíduos se relacionam uns com os outros, o que pode ser observado em diferentes cenários, como política [Tasente 2020, Badawy et al. 2018], saúde [Henneberger et al. 2021] e economia [Seaman et al. 2017]. Modelos de redes sociais, que representam pessoas como nós e suas interações como arestas, trazem grande contribuição para a compreensão do fenômeno de difusão de informação, já que permitem a descrição da heterogeneidade intrínseca das ligações entre os elementos [Pastor-Satorras and Vespignani 2001].

Através do estudo de processos de difusão de informação em redes sociais *online*, é possível obter informações valiosas sobre como a atuação de políticos nessas plataformas podem ajudá-los a vencer eleições [Nguyen 2018], como notícias falsas são espalhadas [Allcott and Gentzkow 2017] e como grupos são coordenados em torno de discussões políticas [Ferreira et al. 2021] e como comportamentos são adotados [Jha and Verma 2024] em relação a produtos de entretenimento, como filmes [Smith et al. 2018] e músicas [Babul et al. 2024, Rosati et al. 2021], que é o foco deste trabalho. A forma como as pessoas descobrem, escutam e compartilham músicas tem sofrido uma profunda transformação causada pela popularização de plataformas de *streaming* e redes sociais voltadas centradas nessa temática [Babul et al. 2024]. Neste trabalho, pretende-se investigar o engajamento de usuários a músicas como um processo de difusão em uma rede social. Particularmente, estamos interessados em compreender como aspectos relacionados ao perfil de consumo de música de usuários e suas conexões com outros usuários em uma rede social refletem sua adoção a novas músicas.

O estudo aqui apresentado é realizado a partir da observação do comportamento de usuários na Last.fm¹, uma plataforma de *streaming* e recomendação musical que armazena os hábitos de escuta de seus usuários a partir da coleta de outras plataformas como Spotify, Apple Music e Deezer. A Last.fm fornece integração social, o que possibilita aos seus usuários acompanhar hábitos musicais de seus amigos e interagir com outros usuários com gosto musical semelhante. Mais especificamente, este trabalho explora músicas da banda Linkin Park (LP), uma banda de rock estadunidense que alcançou sucesso mundial nos anos 2000. Em 2017, a banda enfrentou uma tragédia com a morte do vocalista Chester Bennington, levando a uma pausa em suas atividades. Um dos principais motivos para a adoção da Linkin Park como objeto de estudo se dá pela recente mobilização em torno da banda após o anúncio, em setembro de 2024, de seu retorno com uma nova vocalista, Emily Armstrong, que gerou intensos e controversos debates entre grupos de fãs. O *hype* em torno da banda foi, então, potencializado na ocasião do lançamento do disco “*From Zero*”, em dezembro de 2024, contendo músicas inéditas. Assim, este trabalho aproveita a movimentação sobre a Linkin Park nas redes sociais e concentra suas análises e resultados na investigação da adoção de músicas da banda em um período de 15 semanas em torno do lançamento do disco “*From Zero*”, entre o fim de 2024 e início de 2025.

É possível encontrar trabalhos na literatura que exploram o processo de difusão de informação a partir de informações explícitas sobre o contato entre pessoas para a adoção de uma determinada ideia, como é o caso de [Babul et al. 2024] e [Zhu et al. 2021]. Entretanto, na maior parte dos trabalhos que apresentam estudos sobre processos de difusão em rede ([Mønsted et al. 2017, Traag 2016, Leskovec et al. 2007, Aral and Nicolaides 2017, Cencetti et al. 2023]), não é possível haver uma clareza sobre a forma como ocorre o contato entre os indivíduos, o que pode ser visto como um grande desafio na modelagem do fenômeno. Assim, pode-se partir da hipótese que o processo de difusão pode ser modelado tomando como base uma na observação da atividade de usuários em uma janela temporal, considerando seus perfis de comportamento (hábitos musicais, nesse caso). O processo de adoção de um comportamento por um grupo de usuários pode ser investigado do ponto de vista de análise do fenômeno de propagação da ideia ou informação, mas também do ponto de vista de um modelo de difusão. Modelos

¹ <https://www.last.fm/>

de processos de difusão são, em grande parte, inspirados por modelos epidemiológicos e, de fato, a difusão em redes sociais pode se assemelhar, com frequência, a curvas de contração de doenças contagiosas [Pastor-Satorras and Vespignani 2001], o que justifica, inclusive, a descrição de músicas populares como “virais” [Rosati et al. 2021]. A probabilidade de ativação de nós em redes sociais está associada à velocidade do espalhamento, mas também à própria estrutura topológica da rede, que tem um papel central no fenômeno [Newman 2002, Pastor-Satorras and Vespignani 2001]. É possível apontar, fundamentalmente, duas famílias de modelos de difusão de informação baseados em modelos epidemiológicos, uma baseada em *threshold* [Granovetter 1978] e uma baseada em cascatas de informação [Kempe et al. 2003]. Em uma perspectiva passiva, em modelos baseados em *threshold* um nó é ativado, isto é, adota um comportamento, quando seus vizinhos estão ativos em uma proporção maior que determinado limiar. Em uma perspectiva ativa, modelos baseados em cascata consideram que cada nó ativado tem uma chance, com uma determinada probabilidade, de ativar seus vizinhos. Neste trabalho, o fenômeno de adoção de músicas em uma rede social online é investigado sob o ponto de vista do *Independent Cascade Model* (ICM), baseado em cascatas de informação.

Então, o objetivo principal deste trabalho é investigar o engajamento de usuários a músicas considerando suas interações sociais, que definem uma rede, e seus hábitos musicais, que definem seu perfil. Dessa forma, a seguinte Questão de Pesquisa é levantada para guiar a metodologia aqui apresentada: “*Como o engajamento de usuários a músicas pode ser caracterizado e modelado como processo de difusão de informação?*”. Para construir uma resposta a essa questão central, duas questões de pesquisa auxiliares são levantadas: QP1: “*Como é possível caracterizar o comportamento de usuários em relação à adoção de um comportamento?*”; QP2: “*Quais os papéis dos perfis dos usuários e das suas inter-relações na difusão de um comportamento de engajamento musical?*”

Os experimentos descritos neste trabalho são aplicados, especificamente, a músicas da banda Linkin Park na plataforma Last.fm e, obviamente, a generalização dos resultados apresentados neste trabalho tem sua validade ameaçada por limitações e vieses intrínsecos aos contextos onde as generalizações são realizadas. Porém, é importante apontar que, do ponto de vista metodológico, este trabalho define um *framework* mais geral, que pode ser aplicado a qualquer plataforma que tenha o funcionamento semelhante (onde usuários se seguem e possuem um perfil de comportamento) e a qualquer banda, podendo, inclusive, ser estendida a outros cenários onde seja possível observar a adoção de comportamentos, seja um posicionamento ideologicamente orientado ou o consumo de outros produtos de entretenimento, como filmes e séries.

2. Trabalhos relacionados

A difusão de informação em redes sociais tem sido investigada em diversos contextos, como economia [Seaman et al. 2017], saúde [Henneberger et al. 2021] e política [Ferreira et al. 2021]. Além disso, diversas plataformas são consideradas pelos estudos, explorando a caracterização da dinâmica em redes sociais considerando diferentes abordagens. Por exemplo, Badawy *et al.* tomam como base o Twitter para investigar a propagação de uma ideologia política de usuários, buscando a caracterização da manipulação de seu posicionamento em relação à campanha presidencial de 2016 nos EUA [Badawy et al. 2018]. Tasente considera o Facebook para caracterizar a atuação de *bots* para a manipulação da opinião pública e a rápida difusão de informação não

verificada a respeito do referendo sobre o Brexit [Tasente 2020]. Uma discussão mais aprofundada sobre a dinâmica de notícias falsas é explorada por Allcott e Gentzkow, que investigam sua origem e propagação em redes sociais, de uma forma mais ampla [Allcott and Gentzkow 2017]. Ferreira *et al.* modelam uma rede de co-comentaristas de *posts* no Instagram para caracterizar a evolução das comunidades de usuários e dos tópicos por eles discutidos a respeito de eleições presidenciais no Brasil e na Itália [Ferreira et al. 2021].

A dinâmica social é também estudada por diversos autores do ponto de vista da adoção explícita de um determinado comportamento, como é o caso do trabalho de Aral e Nicolaides, que investiga as relações entre as pessoas em uma rede social de corrida e encontra evidências de contágio social da prática de corrida [Aral and Nicolaides 2017]. Considerando, especificamente o contexto de adoção de músicas, Babul *et al.* analisam como conexões sociais influenciam a descoberta de músicas no Spotify, demonstrando que interações entre usuários afetam recomendações e padrões de engajamento musical [Babul et al. 2024]. Além disso, ele investiga a influência de fatores contextuais, como proximidade social e características da rede, na adoção de novas músicas. Jha *et al.* investigam como os usuários interagem com conteúdo musical em plataformas de redes sociais, como Twitter e Instagram [Jha and Verma 2024]. Os autores analisam padrões de compartilhamento, comentários e curtidas relacionados a músicas, além de explorar como algoritmos de recomendação influenciam o engajamento.

O processo de propagação de adoção de músicas em redes sociais pode também ser estudado a partir da simulação de modelos, como é o caso do trabalho de Rosati *et al.*, que parte da premissa que disseminação de músicas entre as pessoas segue padrões que podem ser descritos por modelos epidemiológicos e utiliza o modelo SIR para caracterizar o engajamento de usuários da plataforma MixRadio [Rosati et al. 2021]. Padrões de comportamento de usuários na Last.fm são investigados por Aljanaki *et al.*, que descrevem a forma como as pessoas consomem música e interagem socialmente na rede [Aljanaki et al. 2010].

3. Materiais e métodos

3.1. Descrição dos dados

A metodologia proposta neste trabalho para a caracterização da adoção de um artista por usuários de uma rede social *online* e a simulação da difusão dessa adoção é genérica o suficiente para tornar sua adaptação e aplicação a outros artistas, contextos temporais e, até mesmo, outras plataformas de rede social. Entretanto, o escopo do trabalho foi mantido em uma rede social *online* específica – Last.fm –, e em um momento específico – as semanas que antecederam e sucederam o lançamento de um novo disco (*From Zero*). Os dados foram coletados utilizando a API oficial da plataforma, que fornece, para um usuário específico, diversas informações, das quais foram utilizadas: faixas e artistas escutados em cada semana e os usuários por ele seguidos. Vale apontar que a API impõe um limite na coleta de 50 usuários seguidos para cada usuário.

Para a coleta, foi definido um conjunto de usuários utilizados como fonte da busca – que neste trabalho serão chamados de *usuários-ego*. A partir de cada *usuário-ego*, foram coletadas as informações de todos os usuários por ele seguido e, para cada um desses, as informações dos seguidos, sucessivamente, até um nível de profundidade

três. Foram definidos como usuários-ego os 250 maiores fãs da banda Linkin Park, como fornecidos pela própria plataforma do Last.fm e este é um critério totalmente arbitrário, escolhido apenas para buscar que os dados envolvessem, de fato, alguma dinâmica sobre a banda. A coleta foi realizada durante 15 semanas, no intervalo entre 01 de setembro e 15 de dezembro de 2024. Ao final, foram coletados os seguintes dados sobre 234000 usuários: as faixas e artistas escutados em cada semana, que determinam o *perfil de consumo musical* de cada usuário, assim como os usuários por eles seguidos, que permitem a construção de uma rede social.

3.2. Modelo de rede

A partir da coleta dos dados sobre usuários do Last.fm com base nos 250 maiores fãs de Linkin Park na plataforma e, a partir deles, uma busca em uma profundidade três, foi construída uma rede social, que serve como objeto principal para a aplicação da metodologia proposta neste trabalho. Essa rede social pode ser representada por um grafo direcionado $G = (V, E)$, onde V é o conjunto de vértices, que representa os usuários da Last.fm, e E é o conjunto de arestas (v_i, v_j) , que representa que um usuário representado pelo vértice v_i segue um usuário representado por v_j . Cada aresta de E é ponderada pela similaridade do perfil de consumo musical dos usuários envolvidos na interação. Foram definidas duas ponderações, w_{ij}^{artistas} e w_{ij}^{faixas} , representando a similaridade entre v_i e v_j em relação ao perfil de artistas e faixas escutados, respectivamente. O peso w_{ij} é calculado através da similaridade de Jaccard, tal que $w_{i,j} = \frac{P_i \cap P_j}{P_i \cup P_j}$, onde P_i é o conjunto de artistas (ou faixas) escutadas pelo usuário representado por v_i na base de dados.

3.3. Modelo de difusão de informação

Considerando a rede social modelada pelas interações dos usuários da Last.fm, como previamente descrita, o modelo *Independent Cascade Model* (ICM) foi aplicado com o objetivo de simular a difusão da adoção de músicas da banda Linkin Park pelos usuários representados. O ICM é amplamente utilizado para simular a difusão de informação, influência ou comportamentos em redes sociais e é baseado na ideia de que a difusão ocorre em passos discretos.

O funcionamento do modelo começa com a ativação de um conjunto inicial de nós chamados de *sementes* da difusão. Em cada passo t , cada nó v_i que foi ativado na rodada anterior $t - 1$ tem uma única chance de ativar cada um de seus vizinhos inativos v_j . A ativação ocorre com probabilidade p_{ij} . Se v_j é ativado, ele muda para o estado ativo e pode tentar ativar seus vizinhos na próxima rodada. Se v_i falha em ativar v_j , nenhuma tentativa futura será feita por v_i para ativar v_j . O processo termina quando não há mais ativações em uma rodada, ou seja, quando nenhum novo nó é ativado.

Tomando como base a rede social representada por G , foi realizada uma inversão na direção das arestas. No modelo de rede original, uma aresta $(v_i, v_j) \in E$ indica que um usuário representado por $v_i \in V$ segue um usuário representado por $v_j \in V$. Porém, no processo de difusão, é mais intuitivo considerar que, se v_i segue v_j , então há uma aresta (v_j, v_i) , já que o ICM tem como premissa que uma aresta (v_j, v_i) indica a influência de v_j sobre v_i . O parâmetro $p_{j,i}$, que representa a probabilidade de v_j ativar v_i , foi ajustado para cada aresta, de acordo como os perfis de consumo musical, considerando a propriedade “artistas” e a propriedade “faixas”, ou seja, proporcionais a w_{ij}^{artistas} e w_{ij}^{faixas} ,

respectivamente. Quatro critérios de escolha de sementes foram considerados, baseados em: usuários que escutaram LP na primeira semana, vértices aleatórios, PageRank e Grau de entrada.

4. Resultados e discussão

Os experimentos conduzidos neste trabalho, e apresentados na presente seção, buscam caracterizar a difusão da adoção de uma banda (Linkin Park) e, especificamente, um álbum (*From Zero*), na rede social *online* Last.fm. Os dados coletados na plataforma Last.fm permitiram a construção de uma rede social direcionada, que representa usuários e relações *seguir*, que na plataforma são descritas como amizades, ponderadas pela similaridade entre perfis de consumo musical de cada par de usuários representados pelas arestas. Essa rede social é o principal objeto, a ser utilizado com o objetivo de responder à Questão de Pesquisa principal apresentada na Seção 1: *Como o engajamento de usuários a músicas pode ser caracterizado e modelado como processo de difusão de informação?*.

Primeiramente, é importante caracterizar os usuários através de suas relações, modeladas como uma rede, indo na direção a uma resposta à QP1 levantada na Seção 1. Cabe ressaltar que a rede social é desconexa, ou seja, há pares de vértices com atingibilidade inviável, independente da direção. Assim, com o objetivo de facilitar a caracterização e a análise do processo de difusão na rede e seguindo uma etapa tradicionalmente utilizada em trabalhos de análise de redes [Newman 2003], foi extraída sua Componente Gigante, i.e., a maior componente conectada, o que resultou em uma sub-rede de tamanho $n = 194600$ vértices e $m = 2549305$, sendo, ainda uma amostra bastante representativa da rede original ($\approx 81\%$ dos vértices). Durante o processo de preparação dos dados notou-se usuários inativos, ou seja, perfis que não apresentavam qualquer tipo de interação ou atividade durante o período analisado ($\approx 19\%$ dos vértices), a exclusão foi necessária para manter a qualidade da análise, já que esses perfis não contribuíam para análise da difusão. Por uma questão de simplicidade, a componente gigante² será tomada como a rede social no restante do trabalho. A rede apresenta, então, um grau médio $\hat{k} = 13.10$ e um Coeficiente de Clustering médio $\hat{c} = 0.08$. Avançando na caracterização da topologia da rede, a Figura 1 apresenta sua distribuição de graus de entrada (Fig. 1(a)) e saída (Fig. 1(b)). É interessante notar que a distribuição de graus de saída (Fig. 1(b)) tem um declínio lento e um *cutoff* em 50, indicando que cada usuário segue, no máximo, outros 50, um viés introduzido no modelo de redes devido a uma limitação da própria API. Mas, mesmo assim, quando observamos a distribuição de graus de entrada (Fig. 1(a)), é possível perceber uma semelhança com distribuições de graus tipicamente observadas em outras redes sociais relatadas na literatura [Newman 2003] e frequentemente utilizadas como *benchmark*³⁴. Também é importante apontar que há uma queda acentuada nas probabilidades observadas para os graus, tanto de entrada, como de saída, indicando que apenas uma pequena quantidade de indivíduos apresenta atividade mais intensa, o que é particularmente relevante no fenômeno de difusão.

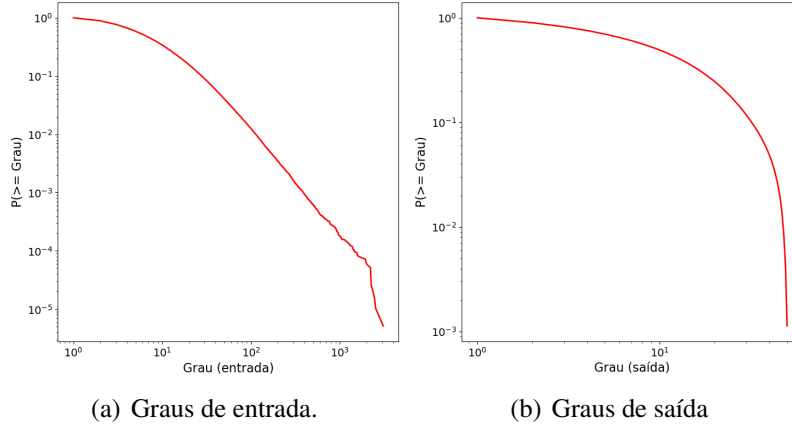
Ainda com o objetivo de investigar o comportamento de usuários com base no comportamento de suas interações na rede social, é possível quantificar a reciprocidade

² Maior subgrafo conexo da rede, em que todos os nós estão interligados por caminhos.

³ <https://snap.stanford.edu/data/>

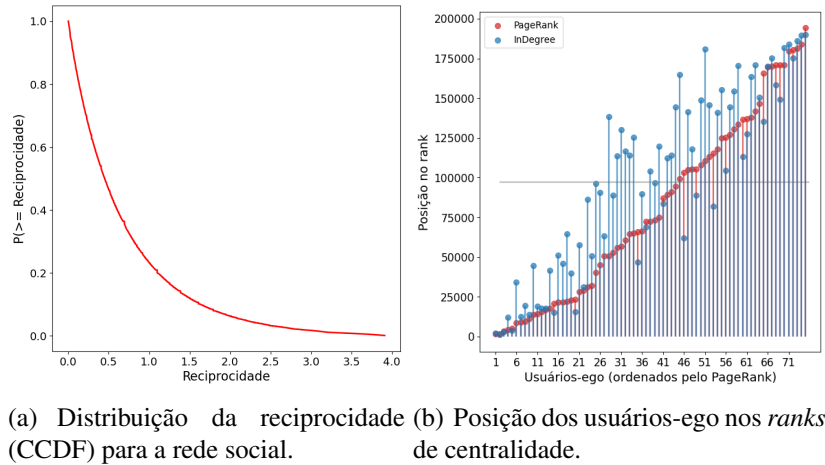
⁴ <https://websites.umich.edu/~mejn/netdata/>

Figura 1. Distribuição de graus (*Complementary Cumulative Distribution Function*) para a rede social.



das relações. Tradicionalmente, a reciprocidade é utilizada como uma forma de quantificar as arestas do tipo (v_j, v_i) que ocorrem em razão de arestas (v_i, v_j) em um grafo direcionado. Considerando a medida de reciprocidade tradicional, foi observada uma reciprocidade global de 0.42. Por outro lado, alguns autores discutem a uma definição de reciprocidade de relações que caracteriza as relações em redes sociais de uma maneira mais refinada [Wang et al. 2011, Kovanen et al. 2010]. Wang *et al.* propõem uma medida de reciprocidade para avaliar o desbalanceamento no relacionamento entre dois usuários $R_{ij} = |\ln(p_{ij}) - \ln(p_{ji})|$, onde p_{ij} é a probabilidade de conexão de um v_i com um v_j [Wang et al. 2011]. A Figura 2(a) apresenta a distribuição da reciprocidade dos usuários na rede social, onde valores mais baixos indicam relações mais balanceadas. É possível

Figura 2. Caracterização topológica da rede social.



perceber que a maior parte das relações apresenta reciprocidade moderada, enquanto poucas relações são extremamente balanceadas ou desbalanceadas. Esses resultados, quando comparados com os apresentados em [Wang et al. 2011], em que os autores também realizam uma análise de características de relações em sociais, revelam uma forte semelhança entre as redes sociais investigadas neste trabalho e as redes investigadas por outros auto-

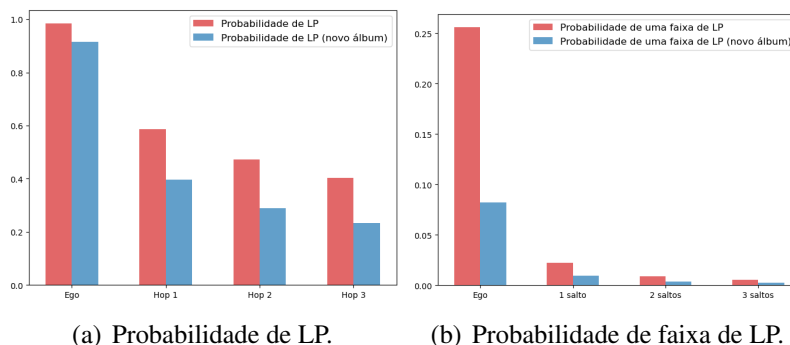
res. Dessa forma, apesar de uma discussão mais aprofundada sobre as causas e o impacto da reciprocidade na rede social estar além do escopo do estudo, os resultados obtidos em relação à reciprocidade representam mais uma forma de validar a metodologia aqui conduzida para a construção de uma rede social a partir de dados de uma plataforma de rede social *online*.

A metodologia proposta neste trabalho toma, como base, um conjunto arbitrário de usuários-ego, especificamente, as 250 pessoas que mais escutaram a banda Linkin Park, um conjunto de usuários fornecido pela própria plataforma do Last.fm. Avançando mais profundamente na QP1 (Seção 1), uma investigação sobre esses usuários foi feita com o objetivo de melhor caracterizar sua centralidade e, assim, compreender sua representatividade e seu papel na topologia da rede. Primeiramente, verificou-se que, dos 250 usuários-ego, apenas 75 estão presentes na componente gigante da rede social original, ou seja, ao filtrar as componentes menores, elimina-se 175 dos usuários-ego da análise. A construção da rede, como proposta neste trabalho, tem uma perspectiva de saída das arestas. Ou seja, são coletados os usuários seguidos por outros (e, sucessivamente, os seguidos por esses). Assim, não há nenhuma garantia, nem expectativa, que os usuários-ego exerçam papel relevante em relação à centralidade topológica na perspectiva de entrada, o que poderia indicar alguma influência sobre seus seguidores na difusão de informação. Nesse sentido, foram calculadas as centralidades de PageRank [Brin and Page 1998] e de grau de entrada dos vértices da rede e, então, foram gerados dois *ranks* em ordem decrescente desses valores. A Figura 2(b) apresenta as posições nos *ranks* calculados a partir do PageRank e do Grau de Entrada (eixo- y) dos 75 usuários-ego ordenados presentes na componente gigante da rede social ordenados pelo PageRank. A análise da Fig. 2(b) permite constatar que, de fato, os usuários-ego não têm relevância topológica de uma forma geral. Ao tomarmos um vértice aleatório na rede, tem-se uma expectativa que ele ocupe uma posição mais próxima ao centro do *rank*. Considerando que a rede social utilizada neste trabalho possui $n = 194600$ vértice, espera-se que um vértice aleatório ocupe, então, uma posição próxima à 97300 no *rank* (representada no gráfico da Fig 2(b) por uma linha cinza horizontal). Entretanto, percebemos que apenas cerca de 60% dos usuários-ego estão em uma posição mais alta que a metade do *rank* baseado no PageRank (e isso não é muito diferente no *rank* baseado em grau). Dessa forma, mesmo sendo tomados como usuários-ego, não há a partir desse experimento, qualquer evidência que os maiores fãs da banda Linkin Park são eficientes em difundir o consumo dessa banda.

De qualquer forma, é razoável considerar que se um usuário v_i segue um usuário v_j , então há uma tendência de semelhança entre os seu perfil de consumo o de v_j . Partindo dessa hipótese, foi realizado um estudo do impacto da topologia da rede social na probabilidade de um usuário adotar a Linkin Park, banda tomada como foco neste trabalho, na janela de tempo observada. Para isso, foram tomados os usuários-ego e os usuários antecessores, em diferentes níveis (*hops*) na rede. Foram considerados, então, os usuários-ego e, seguindo suas arestas de entrada, os vértices a uma distância 1 (1 salto), distância 2 (2 saltos), distância 3 (3 saltos). Com base nesses conjuntos de usuários estratificados, foram realizadas duas medidas. Na primeira, verificou-se a probabilidade dos usuários de cada conjunto adotarem o comportamento de escutar Linkin Park na janela de tempo observada, ou seja, a proporção de usuários que escutou a banda. Na segunda medida, verificou-se a probabilidade dos usuários de cada conjunto escutarem uma faixa de Linkin Park, ou seja, a proporção de músicas da banda entre todas as escutadas pelos usuários

na janela de tempo observada. A Figura 3 apresenta o resultado desse experimento.

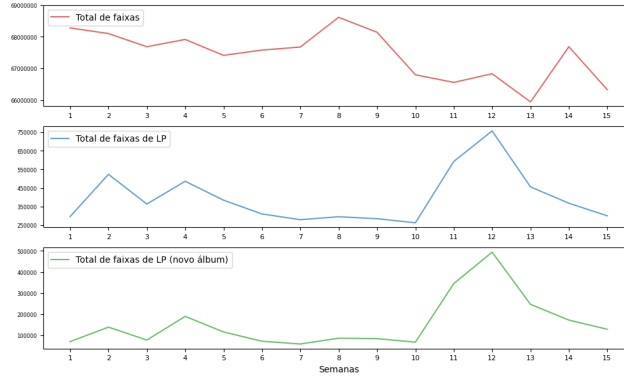
Figura 3. Probabilidade de adoção de um comportamento em razão da distância topológica aos usuários-ego.



De uma forma global, a partir de uma investigação rápida sobre a base de dados, é possível observar que 39% dos usuários escutaram Linkin Park na janela de tempo e 22% escutaram o álbum recém lançado. Ou seja, para um usuário típico, há uma probabilidade de 0.39 que ele tenha escutado Linkin Park (0.22 para o disco recém lançado). Observa-se pela Figura 3 que, de fato, a topologia da rede social tem um papel fundamental na adoção de um comportamento (escutar músicas de uma banda específica, neste caso). Há uma probabilidade de quase 1.00 de que um usuário-ego tenha escutado LP na janela de tempo (Fig. 3(a)) e uma probabilidade próxima de cerca de 0.60 que um usuário que o siga tenha escutado a banda, ou seja, uma probabilidade $\approx 54\%$ maior que a de um usuário aleatório. Mesmo quando afasta-se para uma distância 2 (hop 2), há uma probabilidade de cerca de 0.50 ($\approx 28\%$ maior que a de um usuário aleatório). Ou seja, percebe-se que a topologia exerce um papel importante sobre o perfil de consumo no cenário investigado, mesmo que indiretamente. Uma tendência semelhante pode ser observada para a adoção do disco recém lançado da banda, como também pode ser observado pela Fig. 3(a). Ainda que existente, um impacto inferior da topologia da rede é observado na probabilidade de, ao escutar uma faixa, essa seja de LP, como pode ser observado na Fig. 3(b). Nota-se que há uma queda brusca na probabilidade quando passa-se do ego para usuários que o seguem. Assim, a caracterização dos usuários em relação à adoção da banda Linkin Park, levando em consideração a estrutura da rede, revela um forte indicativo que há um processo complexo que pode ser modelado como um processo de difusão na rede social em relação à adoção da banda na janela de tempo de observação e complementa a resposta à QP1.

Com o objetivo de melhor ilustrar o processo de adoção da banda, a Figura 4 apresenta o número de faixas da banda Linkin Park escutadas pelos usuários presentes na base de dados ao longo das 15 semanas, ajudando a construir a base para compreender o processo de difusão que permite avançar sobre a QP2. Nota-se que a adoção de faixas de LP ao longo das semanas é um fenômeno bastante dinâmico e não apresenta comportamento monotônico, o que coloca um desafio para sua análise como um processo de difusão. Também é possível perceber que as curvas de adoção de faixas de LP se diferem da curva de escuta de faixas em geral, indicando que a tendência de LP não segue meramente a tendência de uso da plataforma Last.fm, mas, de fato, é governada por algum fenômeno

Figura 4. Evolução do número de faixas de LP ao longo das semanas de observação.



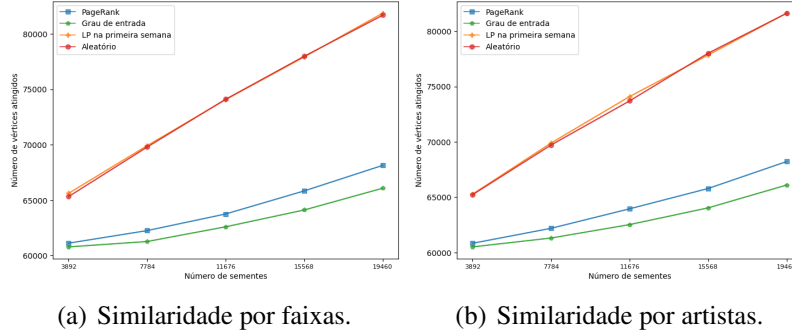
distinto, como é o caso do lançamento de um novo álbum (o que acontece na semana 11, uma semana após o lançamento do álbum)⁵.

Com o objetivo de aprofundar a investigação sobre a adoção de músicas como um processo de difusão, foram realizadas simulações utilizando como base o modelo Independent Cascade Model (ICM) seguindo sua definição padrão, como descrito por seus autores [Kempe et al. 2003]. O modelo foi aplicado à rede social de usuários da plataforma Last.fm, considerando a inversão na direção das arestas. Foram realizadas simulações do ICM considerando uma série de conjuntos de parâmetros. O parâmetro $p_{j,i}$, que representa a probabilidade de v_j ativar v_i , foi ajustado para cada aresta, de acordo como os perfis de consumo musical, considerando a propriedade “artistas” e a propriedade “faixas”. Por se tratar de uma base de dados bastante diversa, os valores de similaridade de Jaccard tendem a ser muito baixos e, por isso, para definição das probabilidades $p_{i,j}$, foram reescalados linearmente para o intervalo $[0.07, 0.13]$, com o objetivo de aproximá-los por valores mais frequentemente utilizados na literatura. Foram considerados conjuntos de sementes de diferentes tamanhos $s_t \times n$, onde s_t está em um intervalo $s = [0.02, 0.10]$, com passo 0.02, considerando os critérios: PageRank, Grau, Aleatório e LP 1^a semana. Para os critérios PageRank e Grau são selecionados os $s_t \times n$ usuários melhor posicionados considerando os *ranks* baseados em PageRank e grau de entrada no grafo original, respectivamente. Para o critério Aleatórios, são selecionados $s_t \times n$ usuários aleatórios como sementes. Para o critério LP 1^a semana são selecionados os $s_t \times n$ vértices que mais escutaram LP na primeira semana. No fim, foram definidos 100 conjuntos de parâmetros, dados pela combinação de todos os parâmetros. Devido à natureza probabilística do ICM, a simulação com cada conjunto de parâmetros foi realizada 10 vezes. A Figura 5 apresenta uma média do número de vértices alcançados pela ICM com cada conjunto de parâmetros. O eixo- x apresenta o número de sementes consideradas e o eixo- y apresenta o número de vértices atingidos.

A partir da Figura 5, primeiramente, podemos perceber que, naturalmente, ao aumentar o número de sementes consideradas, aumenta o número de vértices atingidos, em qualquer cenário, e esse aumento é aproximadamente linear. É possível também notar

⁵Aqui é importante observar que nas semanas que antecederam o lançamento do álbum, houve lançamentos de alguns *singles* de divulgação.

Figura 5. Número de vértices atingidos pelo modelo considerando diferentes métodos de ponderação de probabilidade de transmissão $p_{i,j}$, critérios de escolha e números de semente.



que a difusão acontece com menos intensidade quando são utilizados critérios baseados na centralidade dos vértices (PageRank e Grau). Tomando o experimento realizado de maneira mais ampla, é preciso levar em consideração que a metodologia aqui proposta, baseada em uma rede modelada a partir de dados de uma rede social, tem como premissa que o processo de difusão ocorre de forma que a comunicação entre os usuários acontece através das arestas e sob alguma influência da similaridade dos perfis musicais. Mesmo que, em parte, isso seja verdade, é preciso considerar que há diversos fatores que influenciam na adoção desse comportamento, especialmente em um contexto de grande discussão em torno de um assunto específico, como é, na ocasião, o lançamento do álbum da banda LP. Assim, nesse caso, pode-se argumentar que a topologia da rede tem um efeito limitado sobre o fenômeno da adoção do comportamento de engajamento a músicas no cenário investigado e que a influência entre os usuários pode se dar por outros fatores, como a similaridade de perfil musical, fazendo que os critérios independentes de topologia (LP 1ª semana e Aleatório) sejam capazes de selecionar vértices mais eficientes em difundir o comportamento. Observa-se pela Figura 5 que tomando o critério LP 1ª semana com arestas ponderadas pela similaridade de faixa, cenário em que a difusão teve maior sucesso, o número de vértices atingidos variou de, em média, 65619.9 a 81869.0 (quando o número de sementes são 2% e 10% dos vértices, respectivamente). Esse crescimento de $\approx 25\%$, mesmo moderado, é bastante coerente quando comparado com os resultados de outros trabalhos [Campos et al. 2023, Ling et al. 2023], que utilizam o mesmo modelo de difusão em redes com características semelhantes. Dessa forma, há um conjunto de elementos que permite caracterizar o comportamento dos usuários em relação à adoção de LP (QP1) mas, mais ainda, distinguir a influência dos perfis de consumo musical e suas inter-relações para o processo de difusão (QP2).

Ao final da simulação da difusão, um conjunto de nós é atingido, representando usuários que adotaram o comportamento de escutar à banda. A partir disso, os modelos de difusão construídos são avaliados pela sua capacidade em identificar usuários que, de fato, escutaram LP, observando a base de dados coletada, tratada aqui como *ground truth*. A Tabela 1 apresenta o resultado de precisão e F1-score dos resultados dos critérios de seleção de sementes considerando as ponderações pela similaridade de perfis musicais dos usuários por artistas e faixas. É apresentado apenas o resultado do modelo considerando 10% dos vértices como sementes, o melhor resultado observado em todos os cenários. A

revocação foi próxima de 1.0 em todos os cenários e foi omitida.

	Artistas		Faixas	
	Precisão	F1-score	Precisão	F1-score
PageRank	0.3809	0.5517	0.3813	0.5502
Grau de entrada	0.3823	0.5531	0.3824	0.5532
LP 1ª semana	0.4852	0.6534	0.4845	0.6527
Aleatório	0.3844	0.5554	0.3845	0.5554

Tabela 1. Medida da qualidade dos modelos em identificar usuários que adotaram o comportamento observado.

A partir da Tabela 1 é possível observar que o critério LP 1ª semana é capaz de identificar, com mais eficiência, os vértices que representam os usuários que adotaram o comportamento de escutar músicas da banda ao final da janela de observação. Mais uma vez resultados de difusão relacionados aos critérios baseados em centralidade mostram uma limitação do papel da topologia no processo de propagação da adoção de comportamento no cenário investigado. Mais ainda, percebe-se que, mesmo que as sementes selecionadas pelo critério LP 1ª semana atinjam um número de vértices semelhante ao número de vértices das sementes pelo critério Aleatório, há uma diferença no conjunto de vértices atingidos, que é refletida na distinção da efetividade desses dois critérios em identificar usuários que, de fato, adotaram o comportamento. Assim, mais uma vez, é possível perceber o impacto do perfil de consumo musical dos usuários na sua capacidade de difusão de informação, complementando a compreensão sobre o comportamento de engajamento musical e aprofundando a elucidação da QP2.

5. Conclusões e direções futuras

Este trabalho apresentou uma investigação sobre a difusão da adoção de um comportamento de engajamento musical considerando perfis de usuários em uma plataforma de rede social e suas interações. A metodologia proposta foi aplicada à escuta de músicas da banda Linkin Park (LP) na plataforma Last.fm em um período de 15 semanas, buscando responder à Questão de Pesquisa principal levantada: *“Como o engajamento de usuários a músicas pode ser caracterizado e modelado como processo de difusão de informação?”*

Através da caracterização da rede social formada pelas relações entre os usuários, foi possível perceber que a rede social modelada apresenta características que a aproxima de outras redes sociais frequentemente relatadas na literatura, especialmente quando consideradas as propriedades de distribuição de graus, Coeficiente de Clustering e reciprocidade. Mais ainda, foi possível perceber que a topologia da rede exerce um papel fundamental sobre o fenômeno modelado, já que a proximidade de usuários mais próximos de usuários-ego, que têm um acentuado hábito de escutar LP, apresentam, notadamente, uma maior probabilidade de também adotarem esse comportamento. Foi possível perceber, também, que usuários-ego não exibem, de maneira geral, papel central na topologia da rede. A representação do fenômeno como um processo de difusão utilizando o modelo ICM sobre a rede social modelada permitiu observar que a escolha de sementes baseada na topologia da rede é uma estratégia que não gera cascata de difusão maior do que outros critérios, mesmo aleatório. Também foi possível notar usuários que escutaram a banda na primeira semana tiveram um papel de mais destacado sucesso para a identificação dos usuários que, de fato, adotaram esse comportamento ao longo das semanas.

A metodologia proposta neste trabalho foi aplicada a um caso-de-uso específico e sua generalização demanda a adaptação a outras plataformas e outros contextos, que podem, inclusive, envolver limitações distintas de coleta de dados impostas, o que pode representar uma ameaça à validade dos resultados aqui encontrados. Mesmo assim, pode-se argumentar em favor da validade do presente estudo no sentido da definição de um *framework* para a caracterização e modelagem de qualquer fenômeno de difusão de comportamento em redes sociais, o que pode ser explorado em direções futuras deste trabalho.

Referências

- Al-Garadi, M. A., Varathan, K. D., Ravana, S. D., Ahmed, E., Mujtaba, G., Khan, M. U. S., and Khan, S. U. (2018). Analysis of online social network connections for identification of influential users: Survey and open research issues. *ACM Comput. Surv.*, 51(1).
- Aljanaki, A., Wiering, F., and Veltkamp, R. C. (2010). Mining user behavior in last.fm: Exploring patterns of music consumption and social interaction. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*.
- Allcott, H. and Gentzkow, M. (2017). Social media and fake news in the 2016 election. *The Journal of Economic Perspectives*, 31(2):211–235.
- Aral, S. and Nicolaides, C. (2017). Exercise contagion in a global social network. *Nature Communications*, 8(1):14753.
- Babul, S., Hristova, D., Lima, A., Lambiotte, R., and Beguerisse-Díaz, M. (2024). Link me baby one more time: Social music discovery on spotify.
- Badawy, A., Ferrara, E., and Lerman, K. (2018). Analyzing the digital traces of political manipulation: The 2016 russian interference twitter campaign. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASO-NAM)*, pages 258–265.
- Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117. Archived from the original on 2015-09-27.
- Campos, G., Ribeiro, J., Vieira, V., and Xavier, C. (2023). Estudo do impacto da seleção de sementes baseada em centralidade e em informações de comunidades sobrepostas. In *Anais do XII Brazilian Workshop on Social Network Analysis and Mining*, pages 163–174, Porto Alegre, RS, Brasil. SBC.
- Cencetti, G., Contreras, D. A., Mancastroppa, M., and Barrat, A. (2023). Distinguishing simple and complex contagion processes on networks. *Phys. Rev. Lett.*, 130:247401.
- Ferreira, C. H., Murai, F., Silva, A. P., Almeida, J. M., Trevisan, M., Vassio, L., Mellia, M., and Drago, I. (2021). On the dynamics of political discussions on instagram: A network perspective. *Online Social Networks and Media*, 25:100155.
- Granovetter, M. S. (1978). Threshold models of collective behavior. *The American Journal of Sociology*, 83(6):1420–1443.
- Henneberger, A. K., Mushonga, D. R., and Preston, A. M. (2021). Peer influence and adolescent substance use: A systematic review of dynamic social network research. *Adolescent Research Review*, 6(1):57–73.

- Jha, A. K. and Verma, N. K. (2024). Social media platforms and user engagement: A multi-platform study on one-way firm sustainability communication. *Information Systems Frontiers*, 26(1):177–194.
- Kempe, D., Kleinberg, J., and Éva Tardos (2003). Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, pages 137–146. ACM.
- Kovanen, L., Saramaki, J., and Kaski, K. (2010). Reciprocity of mobile phone calls. *arXiv: Physics and Society*.
- Leskovec, J., Adamic, L. A., and Huberman, B. A. (2007). The dynamics of viral marketing. *ACM Trans. Web*, 1(1):5–es.
- Ling, C., Jiang, J., Wang, J., Thai, M. T., Xue, L., Song, J., Qiu, M., and Zhao, L. (2023). Deep graph representation learning and optimization for influence maximization. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org.
- Mørnsted, B., Sapieżyński, P., Ferrara, E., and Lehmann, S. (2017). Evidence of complex contagion of information in social media: An experiment using twitter bots. *PLOS ONE*, 12(9):1–12.
- Newman, M. E. J. (2002). Spread of epidemic disease on networks. *Phys. Rev. E*, 66:016128.
- Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45(2):167–256.
- Nguyen, J. (2018). Politics and the twitter revolution: A brief literature review and implications for future research. *Social Networking*, 7(4):243–251.
- Pastor-Satorras, R. and Vespignani, A. (2001). Epidemic spreading in scale-free networks. *Physical Review Letters*, 86(14):3200–3203.
- Rosati, D. P., Woolhouse, M. H., Bolker, B. M., and Earn, D. J. D. (2021). Modeling song popularity as a contagious process. *Proceedings of the Royal Society A*, 477(2249):20210457.
- Seaman, C., McQuaid, R., and Pearson, M. (2017). Social networking in family businesses in a local economy. *Local Economy*, 32(5):451–466.
- Smith, M. D., Telang, R., and Catalini, C. (2018). Social media engagement with movies: Analyzing user interactions and sentiment on twitter. *Journal of Marketing Research*.
- Tasente, T. (2020). The #brexit on the facebook pages of the european institutions. *Technium Social Sciences Journal*, 3(1):63–75.
- Traag, V. A. (2016). Complex contagion of campaign donations. *PLOS ONE*, 11(4):1–20.
- Wang, C., Strathman, A., Lizardo, O., Hachen, D., Toroczkai, Z., and Chawla, N. (2011). Weighted reciprocity in human communication networks.
- Zhu, Y., Tang, J., Tang, X., and Chen, L. (2021). Analysis of influence contribution in social advertising. *Proc. VLDB Endow.*, 15(2):348–360.