

Uma Análise sobre a Evolução das Preferências Musicais dos Usuários Utilizando Redes de Similaridade Temporal

Title: On Analyzing User Musical Preferences Evolution through Temporal Similarity Networks

Fabiola S. F. Pereira¹, Cláudio D. G. Linhares¹, Jean R. Ponciano¹,
João Gama², Sandra de Amo¹, Gina M. B. Oliveira¹

¹Faculdade de Computação – Universidade Federal de Uberlândia, Brasil

²INESC TEC – Universidade do Porto, Portugal

{fabiola.pereira, claudiogdl, jean, deamo, gina}@ufu.br, jgama@fep.up.pt

Abstract. *Understanding how user preferences evolve over time is an important personalization task. Traditionally, people's behavior is studied from static or discrete information, assuming that underlying factors such as social influence and personal preferences remain unchanged over time. In this work, we investigate the evolution of user preferences taking into account that such factors are continuous and determinant in the evolution of tastes and choices. We propose to model user preference profiles, in the context of music, through temporal similarity networks. They are able to capture social and temporal characteristics of preferences, taking into account the similarity between users behaviors. We instantiate the proposed model on the social music network This Is My Jam and use temporal measures of centrality and community detection in complex networks. As a result, we detect that the tendency is for artists and similar users to maintain their similarities over time.*

Keywords: *Temporal social networks, similarity networks, user preferences, music social networks.*

Resumo. *Entender como as preferências dos usuários evoluem ao longo do tempo é uma importante tarefa de personalização. Tradicionalmente, o comportamento das pessoas é estudado a partir de informações estáticas ou discretas, assumindo que fatores subjacentes como influência social e preferências pessoais permanecem invariáveis em um longo período de tempo. Neste trabalho, a evolução das preferências dos usuários é investigada levando em consideração que tais fatores são contínuos e determinantes na evolução de gostos e escolhas. Para tanto, é proposta a modelagem de perfis de preferências dos usuários, no contexto de músicas, por meio de redes de similaridade temporal. Elas são capazes de capturar características sociais e temporais das preferências, levando em conta a semelhança entre o comportamento dos usuários. A modelagem proposta foi instanciada sobre a rede social de músicas This Is My Jam. Métricas temporais de centralidade e detecção de comunidades em redes complexas foram utilizadas. Como resultado, foi detectado que a tendência é que artistas e usuários semelhantes mantenham suas similaridades ao longo do tempo.*

Palavras-chave: *Redes sociais temporais, redes de similaridade, preferências dos usuários, redes sociais de músicas*

1. Introdução

Redes sociais *online*, como Facebook, Twitter e redes sociais de músicas, facilitam a construção de relações sociais entre as pessoas que compartilham interesses similares. Conectados uns aos outros, usuários são informados de novas tendências, preferências de consumo e opiniões dos seus amigos, por exemplo. Um processo natural é que as pessoas tendem a mudar seus interesses ao longo do tempo, especialmente em cenários que interagem com uma grande quantidade de itens. Ao mesmo tempo, essas redes sociais crescem e também mudam rapidamente com a adição de novos nós e arestas representando novas interações ou relações na estrutura social. De fato, analisar a estrutura de tais redes pode auxiliar no entendimento de como um usuário modifica seus gostos e costumes ao longo do tempo [Pereira et al. 2018b, Aggarwal and Subbian 2014].

Neste trabalho, o foco está na análise da evolução de preferências dos usuários no domínio de músicas. Entender como essas preferências evoluem é uma das tarefas de personalização mais importantes no contexto de sistemas de recomendação, por exemplo. Quanto mais se conhece sobre o usuário, melhor será a qualidade dos itens recomendados a ele. Os trabalhos em geral propõem observações sobre as variações dos gêneros musicais e artistas com base em atributos como popularidade e periodicidade de audição [Kapoor et al. 2013]. Uma outra linha de pesquisa busca por modelos de aprendizado da evolução das preferências musicais levando em conta a sequência de músicas ouvidas em *playlists* [Moore et al. 2013]. Nenhum desses estudos, entretanto, levam em consideração a influência social que o usuário sofre durante o processo de evolução das suas preferências musicais. De acordo com a psicologia, influência social é quando o comportamento de uma pessoa faz com que outra pessoa mude de opinião ou execute uma ação que, de outro modo, não executaria [Michener 2005].

Tendo como motivação o cenário descrito, neste trabalho é investigada a evolução do gosto musical dos usuários da rede social de músicas *This Is My Jam*¹ (TIMJ), utilizando técnicas de análise de redes sociais com atributos temporais [Holme 2014]. A proposta é representar o perfil atual de preferências musicais de um usuário através de suas semelhanças com os demais usuários, construindo, assim, redes de similaridade temporal – redes nas quais os nós estão ligados com base em alguma semelhança entre eles em um determinado momento.

O objetivo principal do trabalho é verificar se as preferências musicais dos usuários realmente mudaram ao longo do tempo. Para tanto, as seguintes perguntas de pesquisa serão respondidas: (1) artistas semelhantes em relação ao seu público mantêm a semelhança ao longo do tempo? (2) Usuários com mesmo gosto musical evoluem de maneira similar compartilhando das mesmas preferências? A originalidade deste trabalho, que é uma extensão de [Pereira et al. 2018a], está na proposta do uso de redes de similaridade temporal para análise da evolução das preferências dos usuários sobre dados no domínio de música.

Contribuições do trabalho. As principais contribuições podem ser sumarizadas da seguinte maneira: (1) proposta de uma modelagem através de redes de similaridade com atributos temporais para análise da evolução das preferências dos usuários; (2) instanciação da modelagem proposta sobre uma rede social de músicas através de 3 redes

¹<https://www.thisismyjam.com/>

de similaridade temporal – uma rede de similaridade entre artistas, uma rede de similaridade entre usuários com base nos artistas que escutam e uma rede de similaridade entre usuários com base no gênero musical que escutam; (3) uma metodologia para análise de dados modelados na forma de redes com atributos temporais e (4) um conjunto de experimentos executados de acordo com a metodologia proposta, sobre as redes de similaridade temporal instanciadas, e que respondem às questões de pesquisa levantadas acerca da evolução das preferências musicais dos usuários.

Em relação ao artigo [Pereira et al. 2018a], este trabalho estendido incorpora uma nova rede de similaridade baseada em gêneros musicais e novas análises sobre essa rede. Uma metodologia mais acurada para lidar com os vieses da base de dados também é proposta. Além disso, a literatura foi revisada sistematicamente e os conceitos e definições propostos são descritos de maneira detalhada, com muitos exemplos.

Organização do trabalho. Na Seção 2 são apresentados os trabalhos relacionados a esta proposta e suas respectivas contribuições ao estado-da-arte. Na Seção 3 é definido o conceito de redes de similaridade temporal e 3 redes são instanciadas com base no domínio de músicas. Na Seção 4 é proposta uma metodologia para análise das redes instanciadas e, na Seção 5, são apresentados os resultados. A Seção 6 consolida as conclusões obtidas.

2. Trabalhos Correlatos

Este trabalho está relacionado a diferentes áreas de estudo, incluindo evolução de preferências dos usuários, análise de redes temporais e modelagem de redes de similaridade. A originalidade encontra-se na junção destes tópicos aplicada ao domínio de preferências musicais.

2.1. Evolução das Preferências dos Usuários

Segundo [Liu 2015], entender a evolução das preferências dos usuários requer endereçar dois desafios: (i) representação precisa da preferência e (ii) representação precisa de sua dinâmica temporal. De acordo com o campo da psicologia, as probabilidades de escolha das pessoas estão diretamente relacionadas às suas escolhas passadas. Por exemplo, consumidores exibem ou uma lealdade (inércia) ou uma desvalorização (busca de variedade) em relação à última marca comprada [Glanzer 1953]. Seguindo essa linha de raciocínio, então, muitos trabalhos endereçam o problema através sistemas de recomendação temporais [Koren 2009, Xiang et al. 2010, Rafailidis and Nanopoulos 2014], que representam os perfis de usuários por meio de vetores de características, e consideram o histórico desses perfis para prever preferências. São abordagens com foco na predição de futuras preferências sem uma abordagem analítica de aspectos de evolução. Uma desvantagem dessas abordagens é o acaso (super-especialização) que pode levar a modelos enviesados [Debnath et al. 2008].

Existem outros aspectos a serem explorados além da observação do passado para predição. Em outra linha de pesquisa, o foco é na abordagem analítica, buscando a descrição quantitativa e qualitativa da evolução dessas preferências. Esta proposta está mais próxima a essa linha de pesquisa. Os trabalhos variam desde a análise do aspecto evolutivo da *mudança* de preferências ([Moore et al. 2013, Macropol et al. 2013, Kapoor et al. 2013, Lizardo 2018]), da *agregação* de preferências ([Thimm 2013]), até a evolução da *similaridade* entre usuários pelas suas preferências ([Schlitter and Falkowski 2009, Tan et al. 2010]).

Preferências Musicais. Dentre os trabalhos mencionados com foco na abordagem analítica da evolução de preferências, [Schlitter and Falkowski 2009] é o trabalho que mais se assemelha a esta proposta, já que analisa a evolução de preferências musicais dos usuários através da similaridade entre eles. Entretanto, em [Schlitter and Falkowski 2009] o foco é apenas na evolução de comunidades de usuários, representados por um perfil. A diferença é que, neste trabalho, é possível descrever a evolução das preferências musicais sob diferentes perspectivas de similaridade (artistas, usuários e gêneros) e também de generalização (nós e comunidades).

Em [Kapoor et al. 2013] é proposta uma abordagem para predição de mudanças de preferências por meio de um modelo generativo. Os experimentos foram realizados sobre um conjunto de dados que retrata o comportamento musical dos usuários. O trabalho [Moore et al. 2013] desenvolveu um modelo probabilístico que incorpora o tempo para explorar como as preferências musicais de uma população evoluem. A ideia é representar usuários e músicas em um espaço Euclidiano, no qual suas posições mudam ao longo do tempo, demarcando a trajetória da evolução. Em [Calefato et al. 2018] a proposta é explorar a criatividade dos usuários, em caráter de multidão, para incluir diversidade nas comunidade de músicas e mitigar evolução das preferências. Em [Summers and Popp 2015] os autores buscam por padrões de músicas ouvidas de acordo com diferentes contextos. O objetivo é identificar uma relação por exemplo, entre músicas natalinas e a época de Natal, com foco na predição da próxima música a ser tocada em uma *playlist*. São trabalhos de motivação similar a esta proposta, porém não utilizam técnicas de análise de redes temporais.

2.2. Análise de Redes Temporais

Considerando uma rede de larga escala, dinâmica, como aprender os padrões temporais-estruturais de nós e identificar atividades não usuais? Não existe uma metodologia pré-formatada para análise de redes temporais [Pereira et al. 2017]. De acordo com [Aggarwal and Subbian 2014], existem duas maneiras de analisar uma rede que evolui ao longo do tempo: através de métodos de manutenção e através de métodos analíticos. No primeiro, é desejável manter os resultados de um processo de mineração de dados continuamente ao longo do tempo. Por exemplo, em tarefas de predição de *links*, deseje-se manter os modelos atualizados e com boa acurácia à medida que a rede evolui. No segundo, a ideia é quantificar e entender as mudanças que ocorreram na estrutura da rede. São os modelos focados em analisar a evolução, ao invés de apenas se ajustarem a ela. Regras de evolução de grafos que simulam redes do mundo real [Zafarani et al. 2014], bem como regras de densificação e estudos sobre o papel dos nós na dinâmica da evolução da rede [Züfle et al. 2018], são exemplos. Este trabalho adéqua-se a essa segunda maneira de analisar uma rede em evolução – por meio do entendimento das mudanças em sua estrutura.

2.3. Redes de Similaridade

Modelar a similaridade entre objetos na forma de rede não é uma tarefa nova, tendo sido aplicada em diferentes domínios: percepção visual em pinturas [Felício et al. 2016], escrita de artigos científicos [Silva et al. 2015] e imagens [Veit et al. 2017]. A similaridade, por definição, estabelece a força de conexão entre dois objetos de estudo, o que naturalmente pode ser modelado através de uma rede ([Brandao and Moro 2017]).

No domínio de músicas, em [Cano 2004] são analisadas redes de similaridade entre artistas, porém a evolução dessas redes não é levada em conta. E o trabalho [Lizardo 2018] é uma ponte entre a sociologia e ciência de redes com o objetivo de entender as escolhas das pessoas em relação ao gênero das músicas. Por meio de uma rede de gêneros, os autores analisam a trajetória dos usuários na rede, concluindo que gêneros musicais representam nichos de diferentes culturas, e que as pessoas não variam muito suas escolhas. Técnicas de análise de redes sociais temporais [Pereira et al. 2017], entretanto, não foram aplicadas.

3. Redes de Similaridade Temporal

O propósito de construir uma rede de similaridade é conseguir observar como é o comportamento dos nós sob uma perspectiva de relacionamento e influência de uns sobre os outros. Pelo princípio da homofilia [Zafarani et al. 2014], existe uma tendência de grupos de pessoas/itens similares conectarem-se entre si.

Em especial, neste trabalho são consideradas as redes de similaridade temporal, nas quais o *momento* em que as arestas que ligam dois nós similares estão ativas, é um elemento explícito da representação. Ou seja, dois nós podem ser similares em um momento t e no momento seguinte $t + 1$ podem deixar de ser.

Definição 1 (Rede de Similaridade Temporal) *Uma rede de similaridade temporal $G = (V, E)$ é um grafo ponderado não-dirigido, onde E é o conjunto de arestas definidas sobre o conjunto V de nós durante um intervalo de observação $[0, T]$. Uma aresta entre dois nós $u, v \in V$ é representada por uma quádrupla $e = (u, v, t, \sigma)$, onde:*

- $0 \leq t \leq T$ é o momento no qual e está ativa, também chamado de contato;
- $\sigma \in \mathbb{Q}_+^*$ é o peso da aresta e no momento t ;
- e semanticamente indica que u e v são similares no momento t , com um grau de similaridade σ , representando o quão similares eles são.

Exemplo 1 *Como exemplo, considere a rede de similaridade temporal ilustrada na Figura 1. Os rótulos nas arestas indicam seus respectivos pesos σ . Por definição, tem-se que os nós a e b são similares entre si no instante $t = 1$, com grau de similaridade $\sigma = 300$. No instante $t = 2$ eles deixaram de ser similares e, no instante 3, a e b voltaram a ser muito similares ($\sigma = 107$).*

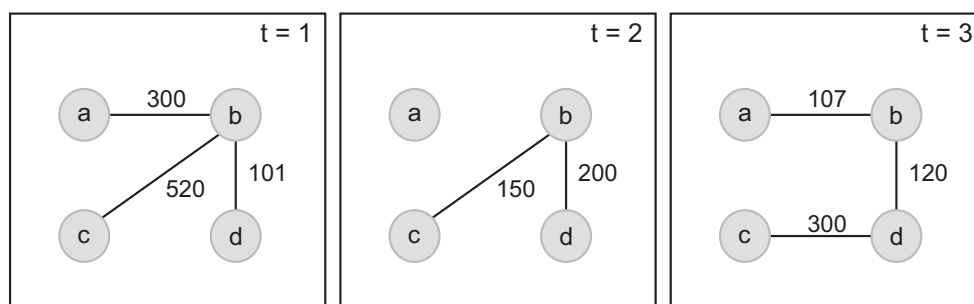


Figura 1. Exemplo de uma rede de similaridade temporal. Se existe uma aresta entre dois nós u e v no instante t , então u e v são similares. Os pesos nas arestas indicam o grau de similaridade σ entre dois nós.

Note que, pela Definição 1, uma rede de similaridade temporal semanticamente representa o relacionamento de nós similares entre si ao longo do tempo. De uma maneira geral, a modelagem de um conjunto de dados na forma de uma rede de similaridade temporal, é realizada a partir dos seguintes passos: primeiro, o objeto sobre o qual deseja-se investigar a similaridade deve ser identificado. Tal objeto será representado por meio de um nó na rede. Em seguida, deve-se determinar a granularidade de tempo a ser considerada, já que trata-se de uma rede temporal (ex.: diária, trimestral, anual). Por fim, o grau de similaridade entre dois objetos deve ser definido, considerando a granularidade de tempo. Esse grau, representado pelo peso σ , indica o quão similar são dois objetos. Opcionalmente, pode-se definir ainda um limiar de similaridade α que servirá como filtro para manutenção das arestas com maior peso ($E = \{(u, v, t, \sigma) | \sigma > \alpha\}$). Tais passos foram utilizados para a modelagem das redes de similaridade temporal sobre o domínio de músicas propostas neste trabalho, descritas a seguir (Seção 3.1).

3.1. Modelagem das Redes de Similaridade Temporal sobre o Domínio de Músicas

As análises foram conduzidas sobre a extinta rede social de músicas *This Is My Jam* (TIMJ). Nela, os usuários podiam compartilhar suas músicas favoritas com seus seguidores. Apenas uma música podia ser compartilhada por vez - o *jam* atual, com duração máxima de uma semana no status dos usuários. Além disso, os usuários podiam curtir o *jam* de outros usuários. A base de dados da TIMJ foi liberada por [Jansson et al. 2015] e contém 219.940 artistas, 132.299 usuários e 2.095.441 *jams*, referente ao período de 26/08/2011 a 26/09/2015.

Originalmente, a base de dados TIMJ não contém a anotação de gênero musical dos *jams*. [Jansson et al. 2015] fez um mapeamento entre a TIMJ e a popular base de dados *Million Song Dataset* (MSD) [Bertin-Mahieux et al. 2011], que é uma base de músicas, largamente utilizada na literatura, com milhões de canções registradas e seus respectivos metadados. Em [Schreiber 2015] foi proposta uma ferramenta capaz de classificar as músicas da base de dados MSD pelo seus gêneros. Assim, para enriquecer a base de dados TIMJ, através da ferramenta de [Schreiber 2015] e do mapeamento de [Jansson et al. 2015], foi possível anotar os gêneros musicais de cada *jam*. Apenas o gênero de maior probabilidade foi atribuído a uma música, não sendo, portanto, uma classificação multi-rótulo. Essa mesma abordagem de classificação de gêneros da TIMJ foi utilizada em [Pereira et al. 2018b].

A seguir são descritas as três redes de similaridade temporal modeladas a partir da base de dados TIMJ sobre as quais as análises acerca da evolução das preferências musicais dos usuários foram conduzidas. Além disso, é apresentada uma discussão acerca do impacto que os parâmetros escolhidos para modelagem podem causar nas análises.

3.1.1. Rede de Similaridade Temporal entre Artistas (*art-art*)

Nessa rede o objeto de investigação é o artista. Períodos trimestrais foram considerados como a granularidade de tempo, totalizando 17 instantes de tempo. Assim, os nós são os artistas e a similaridade ocorre entre artistas compartilhados como *jam* por um usuário em comum dentro de cada instante observado. Quanto mais usuários em comum, maior o peso das arestas. O limiar de similaridade foi empiricamente definido como $\alpha = 20$,

representando que dois artistas são considerados similares se, e somente se, no mínimo 21 usuários compartilharam tanto um artista quanto o outro dentro de 1 trimestre.

Formalmente, tem-se: $G_{art-art} = (V, E)$, sendo V o conjunto de artistas. Para cada aresta $e = (u, v, t, \sigma) \in E$, o grau de similaridade é definido como: $\sigma(e) = |U_u \cap U_v|$, onde U_x é o conjunto de usuários que compartilharam o artista x como *jam* no instante t , para $x \in \{u, v\}$, $t \in [0, 17]$ e $\sigma > 20$.

Como exemplo, considere que a rede de similaridade temporal da Figura 1 seja uma instância de $G_{art-art}$. Considere também que o nó b represente a banda musical *Beatles* e o nó c represente a banda *Radiohead*. 520 usuários tiveram *Beatles* e *Radiohead* como *jams* no primeiro trimestre. No segundo trimestre, esse número diminuiu para 150. E, no terceiro trimestre, as bandas tiveram poucos usuários em comum (no máximo 20), deixando de ser similares em relação ao seu público.

3.1.2. Redes de Similaridade Temporal entre Usuários ($usr-usr_{gen}$, $usr-usr_{art}$)

Nessas redes o objeto de análise é o usuário. Deseja-se relacionar os usuários de acordo com suas similaridades musicais. Assim como nas redes anteriores, foram considerados períodos trimestrais como granularidade de tempo. Com o objetivo de uma análise sob diferentes perspectivas, duas redes foram modeladas – uma que representa a similaridade dos usuários com relação aos artistas que escutam ($usr-usr_{art}$), e outra pela similaridade dos usuários com relação ao gênero musical que escutam ($usr-usr_{gen}$).

Na rede $usr-usr_{art}$ os nós são os usuários e a similaridade ocorre se dois usuários compartilharam o mesmo artista como *jam* dentro do instante observado. O peso das arestas é calculado em função da quantidade de artistas em comum que dois usuários compartilharam ao mesmo tempo. Quanto mais artistas em comum, maior o peso das arestas. O limiar de similaridade empiricamente adotado foi $\alpha = 100$. Formalmente, tem-se: $G_{usr-usr_{art}} = (V, E)$, sendo V o conjunto de usuários. Para cada aresta $e = (u, v, t, \sigma) \in E$, o grau de similaridade é definido como: $\sigma(e) = |A_u \cap A_v|$, onde A_x é o conjunto de artistas compartilhados pelo usuário x como *jam* no instante t , para $x \in \{u, v\}$, $t \in [0, 17]$ e $\sigma > 100$.

Considerando que a rede de similaridade temporal da Figura 1 seja uma instância de $G_{usr-usr_{art}}$, e que os nós a e b sejam, portanto, dois usuários da TIMJ, tem-se: 300 artistas distintos foram *jams* tanto de a quanto de b no primeiro trimestre. Depois, apenas no terceiro trimestre é que a e b voltaram a ter uma similaridade significativa, com 107 artistas em comum sendo compartilhados por eles.

Em uma perspectiva mais genérica, a rede $usr-usr_{gen}$ relaciona os usuários considerando a similaridade em relação aos 15 gêneros musicais da TIMJ. Dois usuários são similares se compartilharam como *jam* músicas com gêneros em comum, levando em conta os fatores diversidade e volume. Formalmente, tem-se: $G_{usr-usr_{gen}} = (V, E)$, sendo V o conjunto de usuários. Para cada aresta $e = (u, v, t, \sigma) \in E$, o grau de similaridade é definido da seguinte maneira: seja $w_x = (g_x^1, g_x^2, g_x^3, \dots, g_x^{15})$, onde cada g_x^i representa a quantidade de vezes que o gênero i foi compartilhado como *jam* pelo usuário x no instante t , para $x \in \{u, v\}$, $t \in [0, 17]$. $\sigma(e) = \sum_{i=1}^{15} \min(g_u^i, g_v^i) / \sum_{i=1}^{15} \max(g_u^i, g_v^i)$, para $\sigma > 0,8$.

Para exemplificar, suponha que um usuário u_1 tenha compartilhado 3 *jams* de *rock*, 2 *jams* de *pop* e 1 *jam* de *jazz* no primeiro trimestre. Já o usuário u_2 compartilhou 3 *jams* de *rock* e 1 *jam* de *blues* nesse mesmo trimestre. Suponha também que existam apenas esses 4 gêneros no domínio e que $w_x = (g_x^{rock}, g_x^{pop}, g_x^{jazz}, g_x^{blues})$. Assim, $w_{u_1} = (3, 2, 1, 0)$, $w_{u_2} = (3, 0, 0, 1)$. O grau de similaridade entre u_1 e u_2 é dado por: $\sigma(e) = (3+0+0+0)/(3+2+1+1) = 0,42$. Note que na rede $usr-usr_{gen}$ o grau de similaridade leva em conta a semelhança entre a quantidade de vezes que o gênero em comum foi compartilhado através das operações de *min* e *max*, refletindo a diversidade. Ao mesmo tempo, o grau de similaridade também leva em conta a quantidade de gêneros que foram compartilhados por um usuário através dos somatórios, representando o fator volume.

3.1.3. Impacto da Parametrização

A Tabela 1 resume as características das redes implícitas analisadas neste trabalho, extraídas da base de dados TIMJ. As redes *art-art* e $usr-usr_{art}$ foram originalmente propostas em [Pereira et al. 2018a]. A rede $usr-usr_{gen}$ é contribuição deste artigo. A escolha da granularidade de tempo como trimestral para todas as redes baseou-se no fato do período ser utilizado também em trabalhos no domínio de músicas [Summers and Popp 2015, Moore et al. 2013]. Em especial, o trabalho [Moore et al. 2013] conclui que 3 meses é o tempo médio que uma pessoa leva para experimentar novas músicas diferentes das habituais.

Tabela 1. Descrição das redes implícitas geradas a partir da rede social de músicas *This Is My Jam*.

Rede	# nós	# arestas	Limiar de Similaridade (α)	Granularidade de Tempo
<i>art-art</i>	1670	3588	20	trimestre
$usr-usr_{art}$	4388	9567	100	trimestre
$usr-usr_{gen}$	5245	53808	0,8	trimestre

Na Tabela 2 estão as propriedades que garantem que as redes geradas são redes do mundo real [Zafarani et al. 2014], com distribuição de graus seguindo a lei de potência, alto coeficiente de clusterização e baixo comprimento médio de caminho entre os nós. O limiar de similaridade (α) foi escolhido empiricamente para cada rede. Quanto menor o limiar, mais arestas existirão na rede e vice-versa. Foi conduzida uma variação empírica de 10 valores diferentes para cada rede, na qual todas as redes geradas continuaram com propriedades do mundo real. Ou seja, variações no limiar não tendem a modificar as propriedades estruturais das redes. O impacto da utilização de um limiar é que quanto mais alto o valor, maior o critério de similaridade exigido. Assim, análises locais do comportamento de um nó em específico podem sofrer variações nas conclusões obtidas. Já análises globais, sobre o comportamento das massas, não.

3.2. Relação entre Redes de Similaridade e Preferências do Usuário

As preferências dos usuários consideradas neste trabalho são as músicas compartilhadas como *jam* na TIMJ, que, por definição, são as músicas favoritas do usuário em um dado instante do tempo. A ideia de investigar as preferências do usuário utilizando similaridade

Tabela 2. Descrição das propriedades que garantem que as redes implícitas de similaridade geradas a partir da rede social de músicas são redes do mundo real. (*lp* denota *lei de potência*).

Propriedade	<i>art-art</i>	<i>usr-usr_{art}</i>	<i>usr-usr_{gen}</i>
distribuição dos graus	lp	lp	lp
coeficiente de clusterização local médio	0,548	0,743	0,742
comprimento médio do caminho	4,57	3,849	3,806

baseia-se no fato de que uma mudança na similaridade é um indício de que os usuários estão mudando seus gostos musicais.

Do ponto de vista dos artistas, os nós densamente conectados indicam que aqueles artistas são ouvidos pelos mesmos usuários e, portanto, são similares em relação ao seu público. Com o passar do tempo, os artistas podem deixar de ser, tornarem-se ou manterem-se similares. Quanto mais mudanças nessa similaridade, mais os usuários estão mudando suas preferências. De fato, a rede *art-art* provê uma visão da evolução das preferências do usuário de uma maneira global. Os usuários podem, por exemplo, mudar suas preferências com base no lançamento de novos artistas ou músicas ou com base em acontecimentos como falecimento ou indicação para prêmios.

Em relação à similaridade entre usuários, a intuição das redes *usr-usr* é que os usuários estão conectados com base na semelhança entre suas preferências musicais. Uma mudança estrutural nessas redes pode indicar que os usuários estão deixando de ter gostos parecidos e tornando-se similares a outros. É uma visão do usuário em relação aos demais membros da comunidade musical, portanto uma evolução de maneira local. Uma pessoa pode mudar suas preferências com base em influências recebidas do meio externo no seu dia-a-dia, deixando de ser similar aos seus pares até então. A rede *usr-usr_{gen}* possibilita analisar essa evolução sob uma perspectiva limitada aos 15 gêneros musicais. Já a rede *usr-usr_{art}* revela detalhadamente as semelhanças entre usuários em relação a cada artista.

4. Metodologia

Este trabalho propõe, então, a análise de redes de similaridade temporal para entender a evolução das preferências dos usuários. Dado o domínio do problema – preferências musicais, as três redes foram modeladas seguindo a semântica de similaridade entre artistas e usuários, conforme apresentado na Seção 3.

Para analisar uma rede temporal, primeiro foram definidas métricas de rede a serem utilizadas. A definição baseia-se na semântica de cada uma em relação ao domínio do problema. Em seguida, foram aplicadas técnicas de visualização de redes para obtenção de *insights* sobre a evolução do comportamento das arestas. Por fim, medidas estatísticas que sumarizam a evolução temporal do comportamento dos nós foram calculadas. A Figura 2 ilustra os passos adotados.

4.1. Definição Semântica de Métricas da Rede

Foram definidas as métricas ponte, influência e versatilidade para análise local dos nós das redes. Cada nó (artista ou usuário) tem o seu comportamento em relação a essas métricas. Também foi analisada a maneira como os nós evoluem em relação às suas vizinhanças ou comunidades. Tais definições são descritas a seguir.

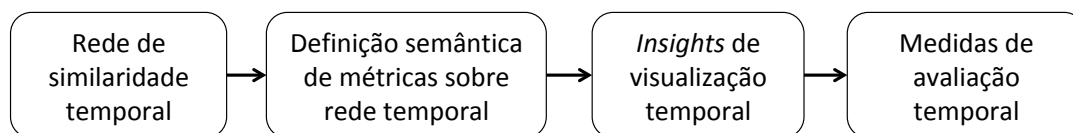


Figura 2. Metodologia adotada para análise de redes de similaridade temporal.

1. **Comunidades:** foi utilizada a métrica *modularity* [Newman 2006] para detecção de comunidades em redes. Ela indica o quanto a estrutura de comunidade encontrada foi criada aleatoriamente. Assim, deseja-se dividir a rede em partições de modo que, se a probabilidade de dois nós se conectarem aleatoriamente for baixa, eles devem estar na mesma partição e vice-versa [Zafarani et al. 2014]. No contexto de similaridade, tem-se que dentro de cada partição encontrada pela métrica estão os nós mais similares entre si.
2. **Ponte:** os “nós ponte” são aqueles que ligam duas ou mais comunidades, mantendo a rede conexa. A métrica utilizada foi a centralidade *betweenness* [Brandes 2001]. Essa medida indica o quão importante o nó é para conectar outros nós. Em redes de similaridade, um nó com alto valor de ponte é aquele que concentra características dos nós das comunidades que ele conecta.
3. **Influência:** um nó influente é aquele importante para os nós da sua vizinhança, que por sua vez, tem alto grau de influência sobre os demais vizinhos, e assim por diante. A métrica *PageRank* [Brin and Page 1998] foi utilizada para detectar a influência de um nó nas redes de similaridade. O *PageRank* de um nó x representa a probabilidade de, a partir de um nó aleatório, chegar a x . Quanto mais alta a probabilidade, maior a influência desse nó na rede. Na similaridade entre artistas, por exemplo, um artista influencia demais artistas parecidos com ele.
4. **Versatilidade:** a versatilidade indica o quão próximo um nó está dos demais nós da rede. A métrica de centralidade *closeness* [Brandes 2001] foi utilizada. Quanto maior o *closeness*, mais similar é o nó em relação aos demais nós da rede, representando portanto um “nó eclético”.

4.2. Insights de Visualização Temporal

Técnicas de visualização da informação podem ser aplicadas para beneficiar a compreensão dos dados e facilitar a identificação de padrões que representem evoluções de uma rede ao longo do tempo, inserindo o usuário no processo e melhorando sua tomada de decisão. Segundo [Linhares et al. 2017], a adoção de uma estratégia de análise visual da rede pode ser o primeiro passo para identificar padrões relevantes muitas vezes inesperados. A ferramenta computacional DyNetVis [Linhares et al. 2017] oferece bons recursos de interação e adota o mapeamento *time-to-space*, que tende a ser melhor para visualizar informações temporais [Beck et al. 2017] e, por isso, foi utilizada para analisar as redes deste trabalho. Além disso, a ferramenta Gephi [Bastian et al. 2009] foi utilizada para obtenção de *insights* de visualização através de layouts estruturais das redes.

4.3. Análise e Avaliação da Evolução das Redes

As comunidades foram avaliadas com base em nós selecionados. Dado um nó $v \in V$, para cada instante de tempo t , sua comunidade C_t^v foi detectada. C_t^v refere-se ao conjunto

de nós pertencentes à mesma comunidade de v no instante t . Depois, a persistência de tal comunidade foi obtida por meio do coeficiente de Jaccard:

$$com(v) = \frac{|C_1^v \cap \dots \cap C_T^v|}{|C_1^v \cup \dots \cup C_T^v|} \quad (1)$$

onde $T = 17$ corresponde à quantidade de instantes de tempo analisada. A persistência de uma comunidade indica, portanto, o quanto uma comunidade varia em relação aos nós que a compõem. Quanto menor a variação, menor a persistência. Quanto mais próximo de 1, maior é a persistência da comunidade de v .

A avaliação da evolução de todos os nós das redes ocorreu em função da variância $var_c(v)$ de cada nó $v \in V$ em relação a uma métrica de centralidade local c durante T instantes de tempo, definida como:

$$var_c(v) = \frac{\sum_{i=1}^T (x_i - \bar{x})^2}{T - 1} \quad (2)$$

onde T é 17, representando a quantidade de instantes de tempo, e x é a distribuição da centralidade c do nó v .

5. Resultados Obtidos

As redes de similaridade entre artistas e usuários foram analisadas de forma que a evolução de cada uma leva à conclusão sobre a evolução das preferências dos usuários de maneira global e local, respectivamente.

5.1. Como foi a evolução dos artistas?

Uma visão estrutural, considerando todo o período de 4 anos, da rede *art-art* é dada na Figura 3. É possível observar que não é uma rede densa e que é composta por diversas comunidades, em sua maioria, pequenas.

A visualização temporal da rede de similaridade entre artistas, mostrada na Figura 4, ilustra a intensidade de comunicação em relação aos 200 nós mais populares ao longo do tempo. No *layout* temporal, os nós são posicionados no eixo vertical e cada instante de tempo no eixo horizontal ilustra as arestas existentes entre os nós naquele momento [Linhares et al. 2017]. É uma perspectiva de atividade temporal na rede. Popularidade [Zafarani et al. 2014], no domínio dos artistas, indica aqueles que mais apareceram como *jam* nos perfis dos usuários considerando todo o período. São os artistas mais curtidos. Pode-se perceber uma interação intensa entre os trimestres 2 e 8 na rede de similaridade, ou seja, uma grande quantidade de artistas populares foram ouvidos pelos mesmos usuários. O artista com maior número de interações foi Radiohead, que ocupa a posição mais central do eixo vertical na Figura 4. Como conclusão, tem-se que os artistas populares foram similares durante a maior parte do período de observação.

Desconsiderando a evolução temporal, a Tabela 3 lista os 10 artistas que se destacam na rede estática, considerando todo o período de aproximadamente 4 anos. É uma visão geral da rede de similaridade *art-art*, que não provê informação sobre a evolução temporal de tais artistas. É interessante observar que os artistas *Radiohead*, *David Bowie*



Figura 3. Visão geral da rede *art-art* considerando todo o período de 4 anos. As cores indicam a comunidade e o tamanho dos nós indica a influência. Alguns dos artistas mais influentes estão em destaque.

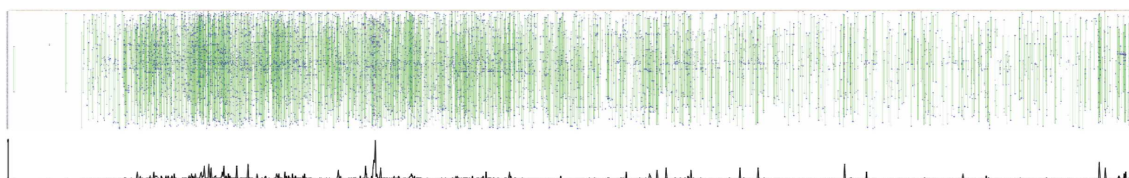


Figura 4. Visualização temporal da rede de similaridade entre artistas considerando os 200 nós mais populares, durante aproximadamente 4 anos com granularidade diária. Visualização gerada pelo software DyNetVis [Linhares et al. 2017].

e *R.E.M.* são aqueles mais centrais em relação a todas as métricas calculadas. Em termos de preferências, a análise estática leva à conclusão de que são os artistas mais preferidos da comunidade de músicas TIMJ.

A Tabela 4 ilustra a evolução dos top-10 artistas com maior variância das métricas locais. A primeira conclusão é que as métricas são complementares e não necessariamente a variância em relação a uma, leva à variância em relação às demais. Os artistas que mais variaram em relação à versatilidade não variaram em relação à influência e ponte. Já entre ponte e influência houve uma sobreposição de 5 artistas dentre os 10 que mais variaram. A variância da centralidade para cada um dos artistas na Tabela 4 mostra que as preferências dos usuários que os escutaram oscilaram ao longo do tempo. Em determinados momentos, *Ramones* e *Pretenders*, por exemplo, foram muito influentes, em outros deixaram de ser.

As comunidades na rede de similaridade entre artistas representam os grupos de artistas que são ouvidos pelos mesmos usuários. Pelos *insights* visuais, dois padrões de evolução de comunidades foram detectados na rede. O primeiro corresponde às comunidades que existiram durante todo o tempo de observação (~ 4 anos), ou seja, uma quantidade representativa de nós que faziam parte de uma comunidade no primeiro ins-

Tabela 3. Top-10 artistas com maiores valores médios das métricas locais. Artistas que se repetiram em todas as métricas estão em destaque.

	Ponte	Influência	Versatilidade
1	Radiohead	Radiohead	Radiohead
2	Art Pepper n Zoot Sims	Irene Dalis	David Bowie
3	Ramones	David Bowie	The Beatles
4	Queen	R.E.M.	Queen
5	Pretenders	Little Jimmy Dickens	The Rolling Stones
6	Rick Derringer	Art Pepper n Zoot Sims	Bob Dylan
7	Pale Saints	Ramones	Nick Cave
8	David Bowie	Beastie Boys	R.E.M.
9	R.E.M.	Joe McPhee Quartet	The Kinks
10	Billie Holiday	Pretenders	Nina Simone

Tabela 4. Top-10 artistas com maior variância para cada uma das métricas locais. Artistas que se repetiram em pelo menos duas métricas estão em destaque.

	Ponte	Influência	Versatilidade
1	Art Pepper n Zoot Sims	Irene Dalis	Ronnie Dyson
2	Ramones	Little Jimmy Dickens	Dire Straits
3	Pretenders	Art Pepper n Zoot Sims	Kanye West
4	Radiohead	Radiohead	Roberta Flack
5	Rick Derringer	Beastie Boys	Aphex Twin
6	Pale Saints	Joe McPhee Quartet	Flower Travellin'band
7	Queen	Pretenders	Refused
8	Billie Holiday	Ricky Nelson	Self
9	The Damned	Ramones	Nicki Minaj
10	R.Seiliog	Rick Derringer	Rocket Ship

tante de tempo permaneceu conectada durante os demais instantes. O segundo refere-se às comunidades que surgiram e desapareceram ao longo da evolução. A Figura 5 ilustra a evolução de duas comunidades durante três trimestres. A primeira, cujo no base é *Radiohead*, possui o comportamento persistente durante a evolução da rede, com $com(Radiohead) = 0,76$. A segunda, com *Ricky Nelson* como artista base, apenas se tornou volumosa no instante $t = 15$, sendo $com(RickyNelson) = 0,13$.

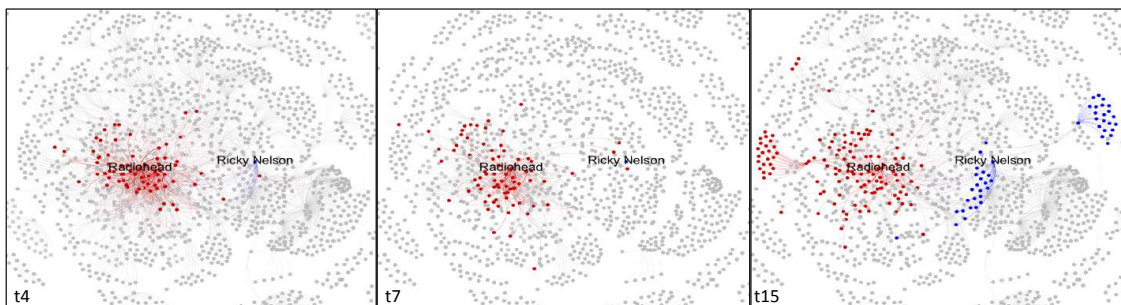


Figura 5. Evolução de duas comunidades a que pertencem *Radiohead* e *Ricky Nelson* em três instantes de tempo diferentes. *Radiohead* pertence a uma comunidade persistente. *Ricky Nelson* pertence a uma comunidade efêmera.

A Figura 6 ilustra a evolução da influência de determinados artistas. É um *heatmap* no qual quanto mais escura a cor, maior a influência. É possível observar que a influência de artistas como *Radiohead* e *Arctic Monkeys* se manteve alta na maior parte do tempo. Entretanto, para os demais artistas, o comportamento detectado é que em geral o pico de influência dura de 2 a 3 trimestres. É uma observação complementar àquela obtida pela análise da evolução das comunidades.

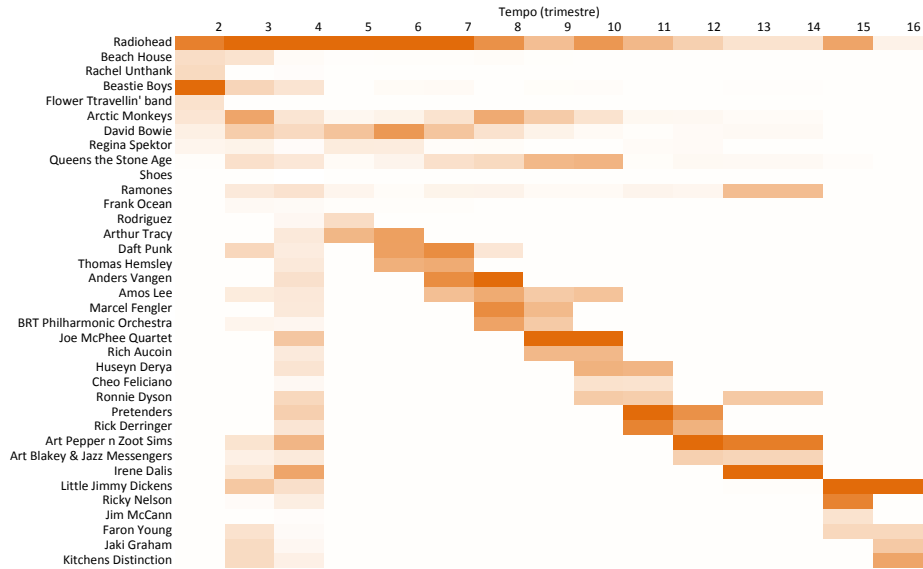


Figura 6. Heatmap baseado na influência dos artistas. Quanto mais escuro, maior a influência. Os artistas da figura são aqueles que em pelo menos um instante de tempo estiveram dentre os 5 mais influentes.

Por fim, o comportamento da variância de todos os nós da rede para cada uma das métricas é mostrado na Figura 7. A média das variâncias tende a zero, ou seja, a maioria dos artistas manteve seus valores de centralidade ao longo do tempo, com baixos índices de variação. A distribuição da quantidade de artistas e suas respectivas variâncias é mostrada nos histogramas. Os top-10 artistas na Tabela 4 estão entre os *outliers* dos *boxplots*.

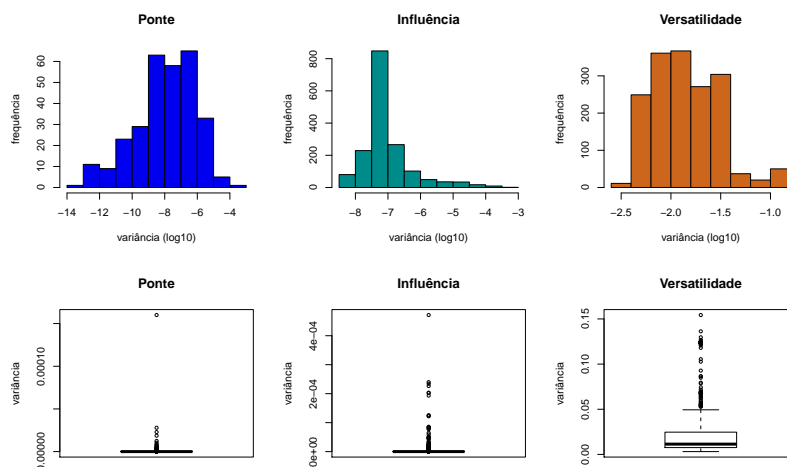


Figura 7. Comportamento da variância dos artistas durante o período observado em relação às métricas de ponte, influência e versatilidade.

A partir das análises realizadas, a primeira pergunta de pesquisa pode então ser respondida: (P1) *artistas semelhantes em relação ao seu público mantêm a semelhança ao longo do tempo?* 97.86% dos artistas mantiveram-se semelhantes ao longo do tempo, tomando como base a variância menor que 0.01 para todas as métricas de centralidade e a persistência das comunidades a que pertencem menor que 0.3. A conclusão é que as

preferências dos usuários não variaram em relação à grande maioria dos artistas. Artistas que eram ouvidos por determinados usuários em comum no início, permaneceram sendo ouvidos pelos mesmos usuários em comum. Os demais 2.14% artistas são *outliers* e tiveram comportamento diferente, com alta variância de centralidade, indicando que seu público oscilou.

5.2. Como foi a evolução dos usuários?

As redes $usr-usr_{art}$ e $usr-usr_{gen}$ têm como principal característica serem muito densas, principalmente a rede por similaridade baseada em gêneros ($usr-usr_{gen}$). Como existem poucos gêneros (15 no total) em comparação com os 132.299 usuários, a chance de dois usuários serem similares em relação ao gênero que escutam é maior do que em relação ao artista específico que escutam. A Figura 8 ilustra a visão estrutural de tais redes. As comunidades encontradas em $usr-usr_{gen}$ revelam que existem de fato 4 grupos de usuários separados pelo gosto ao gênero musical. Esses grupos, entretanto, têm forte interação entre si, não sendo possível perceber a predominância absoluta de certos gêneros em uma comunidade.

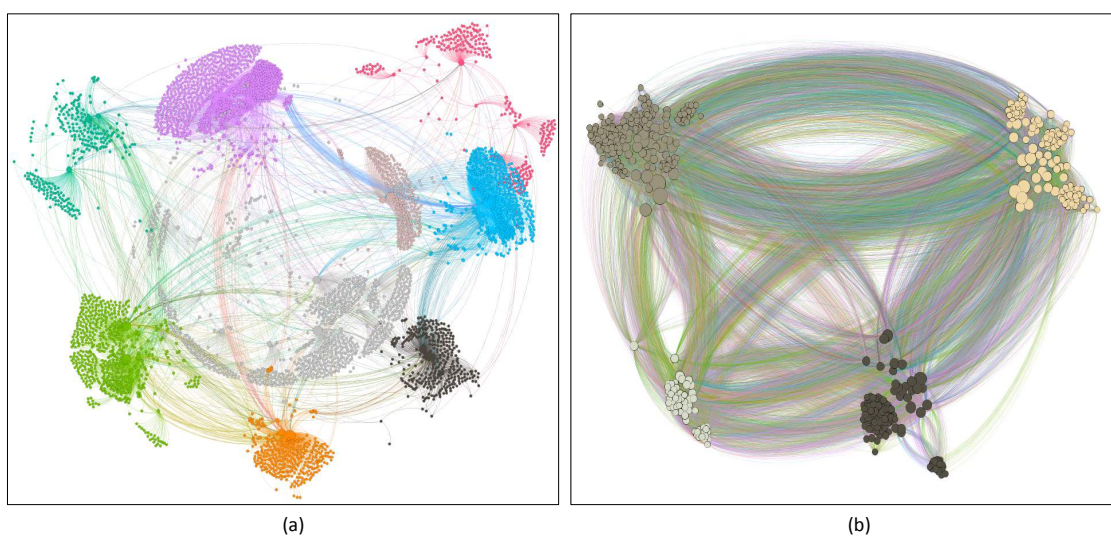


Figura 8. Visão estrutural das redes (a) $usr-usr_{art}$ e (b) $usr-usr_{gen}$ considerando todo o período observado. As cores dos nós indicam as comunidades encontradas e o tamanho dos nós indica a influência. Em (b) as cores das arestas indicam os gêneros musicais.

Ao observar as redes através do layout temporal, destacando a interação entre os usuários mais populares, na Figura 9 é possível perceber uma baixa similaridade entre eles. No contexto de usuários, popularidade foi definida em função do número de seguidores que os usuários possuem. Quanto mais seguidores, mais popular é aquele usuário. A conclusão é que usuários populares formam comunidades de usuários similares a eles e que essas comunidades em geral não se misturam. Pela rede $usr-usr_{gen}$, tem-se que o gênero *rock* (azul) foi predominante durante todo o período de observação. Além do mais, a rede tem um comportamento estrela, evidenciando um usuário com elevado grau de similaridade em relação à maioria dos demais usuários, na maior parte do tempo. É o usuário mais popular e mais versátil da rede TIMJ.

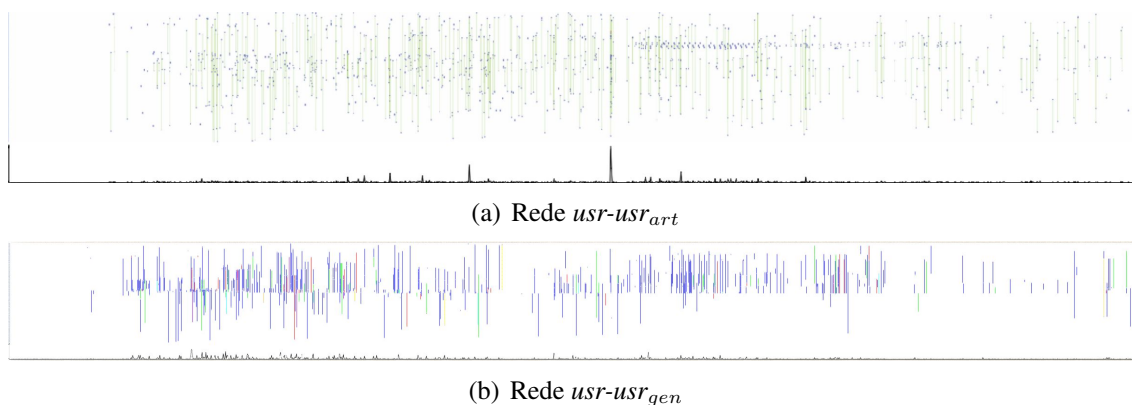


Figura 9. Evolução temporal das redes (a) $usr-usr_{art}$ e (b) $usr-usr_{gen}$ considerando os usuários mais populares. Em (b) as cores representam os gêneros musicais, sendo o gênero de cor azul claramente predominante em todo o período.

Com os *insights* visuais, foi possível identificar que a rede $usr-usr_{gen}$ é menos sujeita a mudanças do que a rede $usr-usr_{art}$. A intuição é que, em geral, as pessoas variam mais entre os artistas que escutam, porém sempre entre gêneros muito parecidos, já que existe claramente um gênero predominante em todo o período.

Para entender a evolução local das preferências de um usuário é necessário observar o comportamento de um dado nó e sua vizinhança local. Utilizando a rede $usr-usr_{gen}$, a evolução da *ego-network* [Zafarani et al. 2014] de um nó u selecionado aleatoriamente, representando um usuário comum da TIMJ, está representada nos *snapshots* da Figura 10. *Ego-network* é uma rede formada a partir de um nó central, chamado *ego*. Esta rede contém também os nós adjacentes ao nó *ego*, chamados de *alters*. A *ego-network* foi obtida considerando todo o período, ou seja, pela rede estática. As influências dos nós também foram calculadas considerando a rede como um todo. Portanto, a evolução das influências não está retratada na Figura 10. Do instante t_4 ao instante t_6 , u mudou suas preferências, tornando-se mais similar a novos usuários. Note que, mesmo sendo uma análise local, a diversidade de gêneros na *ego-network* é alta e não necessariamente os nós são similares apenas em relação a um gênero ou outro. A rede mostra a similaridade considerando volume e variabilidade de gêneros ouvidos. Considerando todo o período observado, tem-se que $com(u) = 0.33$, indicando que a comunidade de u é efêmera.

Em uma perspectiva diferente, a evolução local dos nós também foi observada em $usr-usr_{art}$. A evolução consiste em, a cada trimestre, uma comunidade se destaca com grande volume de interações entre os nós pertencentes a ela, com um nó central sempre em destaque quanto às três métricas investigadas. Tal comportamento pode ser observado pelo *heatmap* na Figura 11, que explora a evolução da métrica de influência para alguns usuários. Os usuários destacados são aqueles que em pelo menos um trimestre estiveram dentre os 5 mais influentes. Diferentemente da rede de artistas, em todos os trimestres, cada um dos 5 usuários mais influentes pertencem a comunidades distintas. A conclusão é que as comunidades nessa rede são efêmeras, ou seja, não existe um grupo de usuários similares que perdura durante todo o período de observação. Além disso, os usuários não estiveram ativos o tempo todo na comunidade. Os picos de interação mostram que a participação na rede social, compartilhando artistas favoritos, é sazonal.

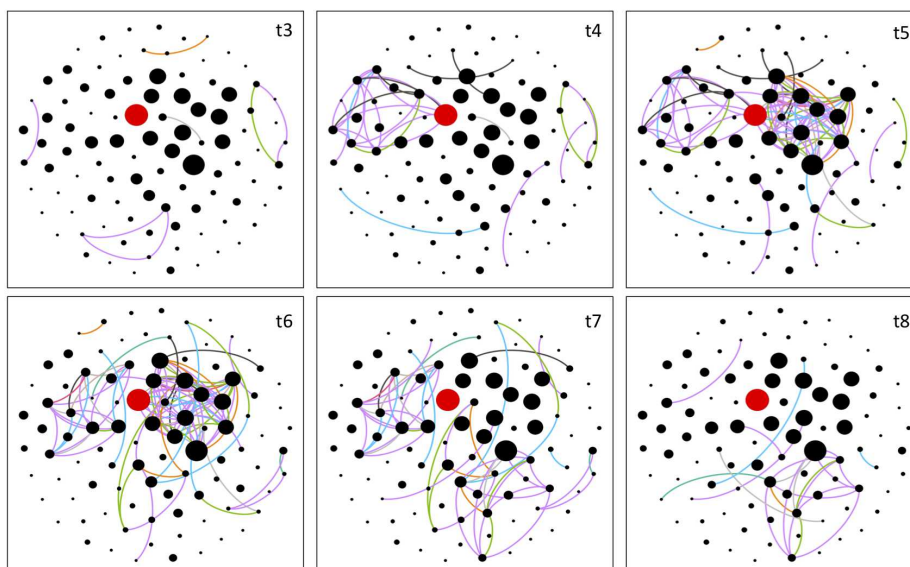


Figura 10. Evolução da *ego-network* do nó em destaque (vermelho). As cores nas arestas representam os gêneros musicais e o tamanho dos nós é proporcional às respectivas influências. Do instante t_4 ao instante t_6 o nó *ego* mudou suas preferências, tornando-se mais similar a novos usuários.

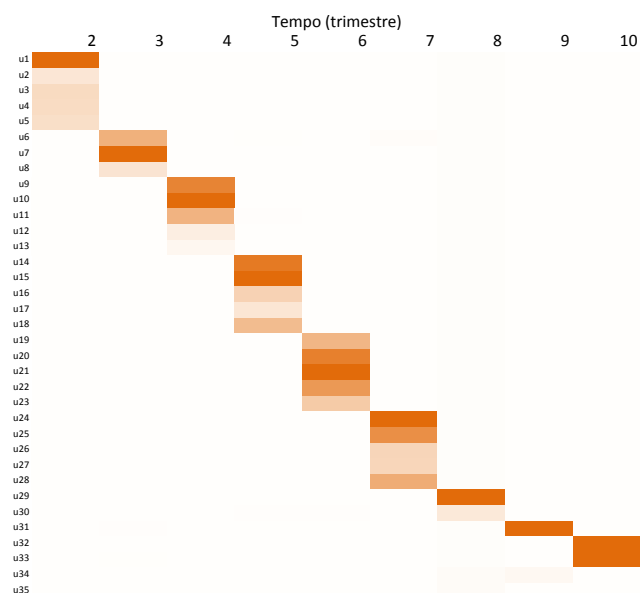


Figura 11. *Heatmap* baseado na influência dos usuários. Quanto mais escuro, maior a influência. Os usuários da figura são aqueles que em pelo menos um trimestre estiveram dentre os 5 mais influentes. Para todos os trimestres, os 5 mais influentes pertencem a comunidades diferentes.

O comportamento da variância entre as centralidades de toda a rede $USR-USR_{art}$ é mostrado na Figura 12. Assim como na rede de artistas, na rede de usuários, a variância da maioria dos nós não foi significativa, tendendo a zero. Quando os usuários estiveram ativos na comunidade, eles se mantiveram similares. Ou seja, os mesmos usuários escutaram os mesmos artistas em comum durante o período de observação.

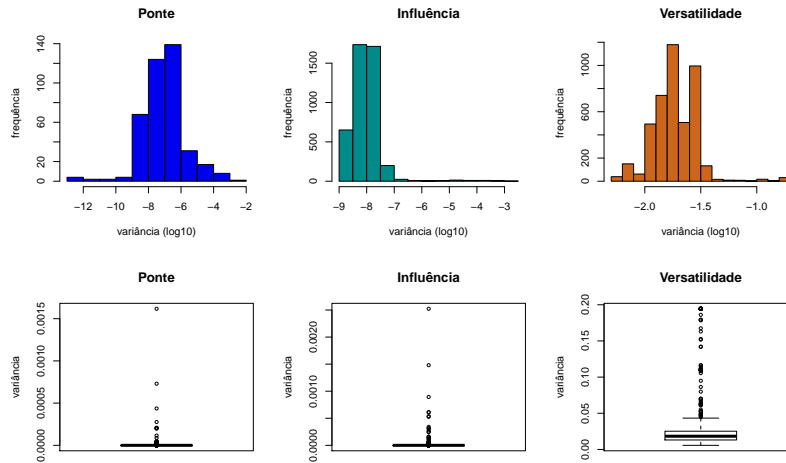


Figura 12. Comportamento da variância dos usuários durante o período observado em relação às métricas de ponte, influência e versatilidade.

Em resposta à segunda pergunta de pesquisa investigada neste trabalho: (*P2*) *usuários com mesmo gosto musical evoluem de maneira similar compartilhando das mesmas preferências?*, tem-se que 98.15% e 97.09% dos usuários mantiveram-se semelhantes ao longo do tempo, tomando como base a variância menor que 0.01 para todas as métricas de centralidade e a persistência das comunidades a que pertencem menor que 0.3, quanto aos artistas e quanto aos gêneros que escutam, respectivamente. A conclusão é que não houve variação nas preferências da maioria dos usuários da rede, uma vez que mantiveram suas similaridades. Os demais usuários tiveram variação significativa (> 0.01 nas três métricas de centralidade), oscilando suas preferências.

5.3. Discussão

As diferentes análises sobre as redes de similaridade temporal mostraram resultados coerentes quanto à evolução de artistas em relação ao público e de usuários em relação às suas preferências. Em geral, a grande maioria dos nós de cada rede não variou suas posições ao longo do tempo, indicando que as semelhanças musicais tendem a se manter. Usuários mantêm a semelhança em relação aos artistas e gêneros que escutam. Assim, se um usuário u tende a mudar de preferência musical, significa que tal mudança é uma tendência do grupo de usuários semelhantes a u como um todo e vice-versa – se o grupo tende a mudar, o usuário u também incorpora a tendência. Da mesma maneira, é o comportamento dos artistas. A tendência é que à medida que um artista possui variação no seu público, todos os artistas semelhantes também possuem essa tendência, mantendo as semelhanças.

Por outro lado, é interessante destacar também o comportamento daqueles nós considerados *outliers*. Os artistas *outliers* são aqueles em plena mudança de público, o que pode ser importante na divulgação de álbuns, por exemplo. Já os usuários *outliers* revelam necessidade de recomendações personalizadas, que evoluem juntamente com suas mudanças. Sistemas que conseguem modelar as mudanças de preferências dos usuários são importantes ferramentas de venda, além de contribuírem para a melhoria da experiência do usuário.

Com os resultados experimentais foi possível mostrar que redes de similaridade

temporal são poderosos instrumentos de análise e modelagem de comportamento baseado na evolução de relações sociais. Elas possibilitaram respostas às perguntas de pesquisa levantadas neste trabalho, inéditas ao estado da arte sobre evolução de preferências musicais.

5.4. Limitações

Toda rede social está sujeita a receber usuários sazonais, ou seja, usuários que acessam a rede esporadicamente [Zafarani et al. 2014]. Em análises pautadas sobre um longo período de observação, como é o caso deste trabalho (aproximadamente 4 anos), os poucos rastros deixados por tais usuários não refletem na íntegra a evolução de seus perfis de comportamento e preferências [Cadilhac et al. 2015]. Não é possível diferenciar, dentre esses usuários, quais acessaram a rede esporadicamente por mudarem suas preferências e quais acessaram esporadicamente simplesmente por indisponibilidade. Usuários sazonais podem ser percebidos nas 3 redes por meio de suas propriedades de distribuição dos graus seguindo a lei de potência (vide Tabela 2). Essa é uma limitação inerente às análises conduzidas, especialmente sobre a evolução dos usuários. De maneira geral, as análises assumiram que a falta de similaridade em um certo período reflete necessariamente as mudanças de preferências.

Os parâmetros de granularidade de tempo trimestral e o limiar de similaridade também são limitações sobre os resultados obtidos. Conforme discutido na Seção 3.1.3, considerar um período anual, por exemplo, para a análise da evolução das preferências, poderia levar a conclusões diferentes. Da mesma maneira, aumentar ou reduzir o limiar de similaridade impacta sobre as conclusões obtidas. A escolha de tais parâmetros foi pautada na literatura correlata em relação ao domínio de músicas, com o intuito de minimizar tal impacto, porém ele ainda existe.

6. Conclusão

Redes sociais são ferramentas que mitigam a construção de relações sociais entre pessoas, bem como o compartilhamento de interesses, preferências e opiniões similares. Assim, analisar como a estrutura de tais redes sociais evolui ao longo tempo auxilia no entendimento de como as pessoas modificam seus costumes e preferências. A modelagem da evolução de preferências é uma importante tarefa de personalização de sistemas de recomendação. Neste trabalho foi conduzida uma análise sobre a evolução das preferências musicais dos usuários de uma rede social de músicas. A proposta foi modelar o domínio através de redes de similaridade temporal. Os experimentos indicaram que a metodologia de análise de redes de similaridade temporal é capaz de detectar e rastrear a evolução das preferências dos usuários em bases de dados grandes e com ruídos. É uma importante vantagem sobre técnicas que não escalam para redes sociais. Foi detectado que 97.86% dos artistas e 97.09% dos usuários mantiveram suas similaridades ao longo do tempo, levando à conclusão de que as preferências musicais dos usuários não mudaram significativamente. As pessoas são ecléticas em relação à diversidade de artistas e gêneros que apreciam, porém tal diversidade é mantida ao longo do tempo.

As características exploradas no domínio de músicas foram artistas e gêneros musicais. Incorporar às músicas metadados como novos álbuns ou nacionalidade dos artistas, por exemplo, são extensões inerentes deste trabalho que podem ajudar a descrever a

evolução das preferências dos usuários através de suas redes de similaridade. Além disso, como futura linha de trabalho, as redes de similaridade temporal são estruturas a serem exploradas para a recomendação temporal de músicas aos usuários.

Agradecimentos

Este trabalho é apoiado pelas agências brasileiras CNPq, CAPES e Fapemig. Os autores agradecem à Microsoft Azure pelo financiamento à pesquisa (*research sponsorship* 65c28dfb-a346-455b-a644-c847ff5ac284).

Referências

- Aggarwal, C. and Subbian, K. (2014). Evolutionary network analysis: a survey. *ACM Computing Surveys*, 47(1):10–36.
- Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks.
- Beck, F., Burch, M., Diehl, S., and Weiskopf, D. (2017). A taxonomy and survey of dynamic graph visualization. *Computer Graphics Forum*, 36(1):133–159.
- Bertin-Mahieux, T., Ellis, D. P., Whitman, B., and Lamere, P. (2011). The million song dataset. *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR)*.
- Brandao, M. and Moro, M. (2017). Tie strength analysis: New metrics and open problems. *VI Brazilian Workshop on Social Network Analysis and Mining*, pages 682–687.
- Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25:163–177.
- Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. In *World Wide Web*, pages 107–117.
- Cadilhac, A., Asher, N., Lascarides, A., and Benamara, F. (2015). Preference change. *Journal of Logic, Language and Information*, 24(3):267–288.
- Calefato, F., Iaffaldano, G., Lanubile, F., and Maiorano, F. (2018). Investigating crowd creativity in online music communities. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW):27:1–27:21.
- Cano, P. (2004). The emergence of complex network patterns in music artist networks. In *Int. Symposium on Music Information Retrieval*, pages 466–469.
- Debnath, S., Ganguly, N., and Mitra, P. (2008). Feature weighting in content based recommendation system using social network analysis. In *Proceedings of the 17th International Conference on World Wide Web, WWW '08*, pages 1041–1042.
- Felício, C. Z., de Almeida, C. M. M., Alves, G., Pereira, F. S. F., Paixão, K. V. R., and de Amo, S. (2016). *29th Canadian Conference on Artificial Intelligence*, chapter Visual Perception Similarities to Improve the Quality of User Cold Start Recommendations.
- Glanzer, M. (1953). Stimulus satiation: An explanation of spontaneous alternation and related phenomena. *Psychological Review*, 60(4):257–268.
- Holme, P. (2014). Analyzing temporal networks in social media. *Proceedings of the IEEE*, 102(12):1922–1933.

- Jansson, A., Raffel, C., and Weyde, T. (2015). This is my jam – data dump. *16th Int. Society for Music Information Retrieval Conference*.
- Kapoor, K., Srivastava, N., Srivastava, J., and Schrater, P. (2013). Measuring spontaneous devaluations in user preferences. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13, pages 1061–1069.
- Koren, Y. (2009). Collaborative filtering with temporal dynamics. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, pages 447–456.
- Linhares, C. D. G., Travençolo, B. A. N., Paiva, J. G. S., and Rocha, L. E. C. (2017). Dynetvis: A system for visualization of dynamic networks. In *SAC*, pages 187–194.
- Liu, X. (2015). Modeling users' dynamic preference for personalized recommendation. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI'15)*, pages 1785–1791.
- Lizardo, O. (2018). The mutual specification of genres and audiences: Reflective two-mode centralities in person-to-culture data. *Poetics*, 68:52 – 71.
- Macropol, K., Bogdanov, P., Singh, A. K., Petzold, L., and Yan, X. (2013). I act, therefore i judge: Network sentiment dynamics based on user activity change. In *Int. Conf. on Advances in Social Networks Analysis and Mining*, ASONAM '13, pages 396–402.
- Michener, H. A. (2005). *Psicologia Social*. Cengage Learning.
- Moore, J., Chen, S., Turnbull, D., and Joachims, T. (2013). Taste over time: the temporal dynamics of user preferences. In *Int. Society for Music Information Retrieval Conf.*
- Newman, M. E. (2006). Modularity and community structure in networks. *Proc Natl Acad Sci U S A*, 103(23):8577–8582.
- Pereira, F. S. F., Gama, J., and Oliveira, G. (2017). *Knowledge Discovery from Temporal Social Networks*, pages 1–19. Information Management Techniques.
- Pereira, F. S. F., Linhares, C. D. G., Ponciano, J. R., Gama, J., Amo, S., and Oliveira, G. M. B. (2018a). That's my jam! a temporal analysis over user preferences evolution in a social music network. *VII Brazilian Workshop on Social Network Analysis and Mining*.
- Pereira, F. S. F., Tabassum, S., Gama, J., Amo, S., and Oliveira, G. M. B. (2018b). On analyzing user preference dynamics with temporal social networks. *Machine Learning*, 107(11):1745–1773.
- Rafailidis, D. and Nanopoulos, A. (2014). Modeling the dynamics of user preferences in coupled tensor factorization. In *Proceedings of the 8th ACM Conference on Recommender systems*, pages 321–324. ACM.
- Schlitter, N. and Falkowski, T. (2009). Mining the dynamics of music preferences from a social networking site. In *Social Network Analysis and Mining, 2009. ASONAM '09. International Conference on Advances in*, pages 243–248.

- Schreiber, H. (2015). Improving genre annotations for the million song dataset. In *Proceedings of the 16th International Society for Music Information Retrieval Conference, ISMIR*, pages 241–247.
- Silva, V., Sampaio, F., and Oliveira, J. (2015). Temporal analysis of co-authorship networks: A study on the interactions of authors in the brazilian journal of computing in education. *Brazilian Journal of Computers in Education (RBIE)*, 23(2).
- Summers, C. and Popp, P. (2015). Temporal music context identification with user listening data. *16th International Society for Music Information Retrieval Conference*.
- Tan, C., Tang, J., Sun, J., Lin, Q., and Wang, F. (2010). Social action tracking via noise tolerant time-varying factor graphs. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '10*, pages 1049–1058.
- Thimm, M. (2013). Dynamic preference aggregation under preference changes. In *Proceedings of the Fourth Workshop on Dynamics of Knowledge and Belief (DKB'13)*.
- Veit, A., Belongie, S., and Karaletsos, T. (2017). Conditional similarity networks.
- Xiang, L., Yuan, Q., Zhao, S., Chen, L., Zhang, X., Yang, Q., and Sun, J. (2010). Temporal recommendation on graphs via long- and short-term preference fusion. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '10*, pages 723–732, New York, NY, USA. ACM.
- Zafarani, R., Abbasi, M. A., and Liu, H. (2014). *Social Media Mining: An Introduction*. Cambridge University Press, New York, NY, USA.
- Züfle, A., Renz, M., Emrich, T., and Franzke, M. (2018). Pattern search in temporal social networks. In *EDBT*.